Testimony by Prof. Reinhard Laubenbacher
Professor, Virginia Bioinformatics Institute and Department of Mathematics
Virginia Polytechnic Institute and State University
and
Vice President for Science Policy, Society for Industrial and Applied Mathematics (SIAM)

Testimony on 21st Century Biology
before the
Subcommittee on Research and Science Education
Committee on Science and Technology
United States House of Representatives, Washington, DC

June 29, 2010

My name is Reinhard Laubenbacher and I am a professor at the Virginia Bioinformatics Institute, where I lead the Applied Discrete Mathematics Group and am the Director for Education and Outreach. I am also a professor of mathematics at the Virginia Polytechnic Institute and State University and an adjunct professor in the Cancer Biology Department at the Wake Forest University School of Medicine.

Since 2009 I have served as Vice President for Science Policy for the Society for Industrial and Applied Mathematics (SIAM). SIAM is a community of approximately 13,000 applied and computational mathematicians, computer scientists, numerical analysts, engineers, statisticians, and mathematics educators who work in academia, government, and industry. While SIAM members come from many different disciplines, we have a common interest in applying mathematics in partnership with computational science towards solving real-world problems.

In my invitation to testify on the New Biology, the Subcommittee raised questions in three areas, and I have organized my testimony accordingly into three sections:

- Research to Address Grand Challenges and Areas of Scientific Opportunity
- Interdisciplinary Collaborations – Culture and Cross-Agency Coordination
- Workforce – Education and Training

In each of these sections, I offer observations from my experiences at the interface of mathematics and biology and specific comments and recommendations about National Science Foundation (NSF) programs. Specifically, the testimony highlights

- ways in which mathematical and computational research will contribute to New Biology research to tackle societal challenges in food, energy, the environment, and health;
- mechanisms for support of research at the interface between mathematical and life sciences, and examples of successful programs in this area;
- lessons learned on the integration of cultures to enable interdisciplinary research; and
- recommendations for ways to enhance graduate and undergraduate education to prepare students to conduct research in the New Biology.

I note that many of the descriptions of research opportunities and the recommendations in this testimony reflect discussion within SIAM on the opportunities interface between the mathematical and computational sciences and the life sciences, as reflected in a white paper SIAM has produced in this area.[1]

RESEARCH TO ADDRESS GRAND CHALLENGES, AREAS OF SCIENTIFIC OPPORTUNITY

First Set of Questions from the Committee: In your opinion, what is the future of research in the biological sciences and what role does research at the intersection of biology and mathematics hold for addressing grand challenges in energy, the environment, agriculture, materials, and manufacturing? What computational tools still need to be developed? Are there promising research opportunities that are not being adequately addressed? Is the National Science Foundation playing an effective role in fostering research at the intersection of the physical sciences, engineering, and the biological sciences? If not, what recommendations would you offer?

The 2009 National Research Council report "A New Biology for the 21st Century: Ensuring the United States Leads the Coming Biology Revolution"[2] proposes a national initiative to promote the New Biology that focuses on problem-centric, interdisciplinary research in the life sciences to solve societal challenges in Health, Food, Energy, and Environment. A central finding of the report is that new information technologies and sciences will be essential to achieving the New Biology and meeting these challenges. Biology has become a highly technology driven, fast moving science. New technologies typically produce new data types and larger volumes of data, and allow that data to be generated more cheaply. At the same time, the expertise, tools, and time needed to analyze that data, to turn it from numbers into knowledge and understanding, is becoming more complex and more expensive. For example, while the cost of sequencing a person's genome is moving toward the $100 level, the cost of extracting information from the sequence that is meaningful for that person's health is likely in the $1 million range. **So the real bottleneck in biology is already shifting toward data analysis. Breakthroughs in mathematics, statistics, and the computational sciences will be necessary to assure that data analysis can keep up with data generation.**

For each challenge area, the report outlines how biology can contribute directly and which research and technological needs must be met in order to do so. **In each area, new approaches to information analysis, data, and modeling will be needed to advance our understanding of the natural world,** as biology develops as a predictive science**.**

Food: In order to help ensure a sustainable and responsibly grown food supply, particularly in light of the changing global climate, one of the challenges is to understand and quantify how plants grow and interact with their environment. This involves characterizing the relationship between the genotype and phenotype of organisms, a fundamental problem in biology. **At the genome level biology is essentially digital, and genetic sequence information is translated into dazzlingly complex interacting networks of genes, proteins, and metabolites, making up cellular function**. Cells organize into tissues, which, in turn form the whole plant.

---

1. The SIAM white paper on "Mathematics: An Enabling Technology for the New Biology" is available on line at http://www.siam.org/about/science/pdf/math_biology.pdf.
2. National Research Council, *A New Biology for the 21st Century: Ensuring the United States Leads the Coming Biology Revolution* (2009), http://dels.nas.edu/Report/Biology-21st/12764.

Functioning of the cellular networks is directly influenced by features of the environment the plant finds itself in, such as climate, resource availability, and microbial communities.

Environment: In order to sustain ecosystem functions in the face of rapid change, we need to be able to monitor multiple heterogeneous variables spanning a range of temporal and spatial scales. The **vast amount of data so collected needs to be integrated and used to construct unifying mathematical models that help guide environmental policy, and have the predictive capability to assess consequences of informed intervention**. Here too, the models need to integrate interconnected networks and systems of complex systems at vastly different scales, all affected by a common environment.

Energy: In order to expand sustainable alternatives to fossil fuels, new approaches beyond ethanol derived from corn must be developed. Microbial biocatalysis, for example, is a promising direction. In order to make it a reality, solving the genotype-phenotype problem will lead to the capability to engineer microbes from standard DNA modules that perform a specified metabolic function. Another promising approach is to engineer plants with molecular networks that produce more leaves and fruit without using additional fertilizer, thereby increasing energy production through photosynthesis. **With predictive models of the intertwined gene, protein, and metabolic networks, it becomes feasible to engineer and optimize the organism for efficient biofuel production**.

Health: To make a transformational contribution to human health, solution of the genotype-phenotype problem will contribute to **integrating genomics information with complex genetic, protein, and metabolic networks, on up to the tissue and organism levels, all of which react to the external environment**. In fact, environmental influences are known to play a very important role in several important diseases, such as cancer and neurological disorders.

The importance of developing better modeling, computational, statistical, and analytical tools to enable a better understanding of biological systems and detailed discussion of the potential impact and key problems are also described in the 2005 National Research Council report "Mathematics and 21st Century Biology."[3] We are approaching a time when gathering the data necessary to truly begin to comprehend complex life as a whole system will be possible. This will be done through consolidating the ever-increasing amounts and types of available information at an ever-increasing level of completeness and granularity. The development of mathematical and computational tools to use this information in sophisticated models should be a priority. To date, exploiting modeling in biology has led to progress on understanding small pieces of large complex systems. But for the biological sciences to bring their full potential to bear on solving the most challenging problems humankind faces in the 21st century, we must now turn our attention to the comprehension of whole systems, and the mathematical and computational sciences are a key enabling technology in this quest.

3. *Mathematics and 21st Century Biology* (2005) is available at http://www.nap.edu/catalog.php?record_id=11315.

Common Themes from Challenges in New Biology Report

Three common themes emerge from the challenges described in the report.

1. **All four challenges require the construction and analysis of predictive mathematical models of large, nonlinear dynamic networks that span several spatial and temporal scales.** Understanding and manipulating these systems will require large, multi-scale, nonlinear, and hybrid models. Existing simulation and analysis tools for such models are in their infancy, or nonexistent in some cases. For instance, an increasingly popular modeling paradigm for complex networks in fields ranging from molecular biology to ecology is agent-based modeling, which captures the important feature of many complex systems that global behavior emerges from local interactions. Very few analysis tools exist for such models. For many applications it is desirable to use models to predict how interventions on one level will impact biological systems on other levels, such as in the development of therapeutics. This process requires control approaches, but for the systems at the heart of the New Biology challenge areas, it is sometimes difficult or impossible to apply existing control theoretic approaches.

2. In all problem areas **high performance computation will play a crucial role, from simulating complex multi-scale models to analyzing sequence data**, e.g., multiple sequence alignment. **This will require new breakthroughs in algorithm development**, since we cannot expect significant increases in clock speed due to silicon technology. Performance improvements in computation will come from more cores on a chip. This means significant changes in algorithms to take advantage of parallelism on the chip as well as parallelism between computational nodes comprised of multiple chips. In order to achieve high rates of performance, algorithms that minimize data movement, possibly at the expense of doing additional computations, will be the most efficient. Algorithm developers will need to take these facts into account as they develop multi-scale, multi-physics algorithms.

It is also important to mention that **the speedup in scientific computation achieved over the last four or five decades owes more to the development of new numerical algorithms than to hardware improvements.** Several reports have documented the ways in which the contribution of algorithms has surpassed the improvements due to better technology (Moore's Law),[4] but the impact from both has been critical. Together, hardware and mathematical improvements account for an increase in the speed at which we are able to perform the calculations to model important systems, such as in numerical weather prediction, by a factor of roughly 10,000,000 in the period between the 1960s and the 1990s.

3. In all four challenge areas we face **ever-growing data volumes**, from DNA sequence data to satellite surveillance data. As an example, the amount of DNA sequence data stored in GenBank, a data repository maintained by the NIH, has grown by a factor of 100,000 over the past 25 years. Currently, there are over 150 million genetic sequences stored in this publicly-available database. **Genetic data** like this, **and the many other types of data generated by the application of new imaging tools and other technologies to biological systems, need to be**

4. See, for example, Figure 5, page 53 of *Computational Science: Ensuring America's Competitiveness*, a 2005 report to the President of the United States from the President's Information Technology Advisory Committee (PITAC). See also Figure 13, page 32 of the DOE Office of Science report *A science-based case for large-scale simulation*, 2003.

stored in databases that are easily accessible, organized, and searchable, requiring increasingly sophisticated and scalable data mining algorithms. In addition, the data from heterogeneous sources need to be integrated, within databases as well as within models. Once accessible in databases, the typically high dimensional data sets need to be analyzed using statistical methods. In order to meet these challenges, new tools from multivariate statistics and discrete mathematics are needed, in particular graph theory and combinatorics.

Biology to Inform Mathematical Research

As happened with physics in the last century, we can expect that an increasingly strong feedback loop will develop between biology and the computational disciplines that now serve as tools, such as mathematics, statistics, computer science, and engineering. For instance, the National Science Foundation is already capitalizing on this feedback with its program "Quantum and Biologically Inspired Computing." We mention here two more examples.

It is well appreciated that the human immune system has important lessons to teach us about computer security. But the immune system is also a vast distributed information-processing network that adapts to ever-changing tasks. Once we understand its design principles well enough to build mathematical models capturing its key capabilities we will be able to transfer these principles to engineered networks. The immune system's complexity and the multiple spatial and temporal scales involved offer several mathematical and computational challenges that can only be overcome by fundamental breakthroughs in these fields.

As another example, it is observed frequently by experimentalists that after engineering an organism with a gene deletion, even an apparently essential one, its phenotype remains unchanged. That is, the organism is robust to many such changes and can remodel its molecular networks after a change in its genome to maintain function. The underlying fundamental problem of understanding the genotype-phenotype relationship is mirrored by the analogous mathematical problem, namely understanding the relationship between the structure of a dynamical system and its resulting dynamics. This problem is still largely unsolved and poorly understood. Biological insights about the sources of this robustness in organisms can help generate hypotheses about solutions to the corresponding mathematical problem in dynamical systems. In turn, these solutions can be applied to better understand and control other complex systems such as the power grid and computer networks.

Recommendations — Research Areas

This analysis makes clear that mathematics is indeed an important enabling technology for the New Biology. **We recommend that any funding programs related to the New Biology initiative provide support for mathematical research related to the problems identified above in the following areas:**

1. Complex networks, both in the graph-theoretic sense and in the dynamical systems sense.
2. Multi-scale modeling and simulation, including computational science research to enable new approaches.
3. Systems of partial differential equations.
4. Algorithms for high performance computation.

5. Algorithms for new multi-core computer architectures.
6. Multivariate statistics.
7. Dynamical systems.
8. Hybrid models.
9. Control theory.
10. Combinatorics and graph theory.
11. Data mining algorithms.
12. New methodologies for modeling complex stochastic biological systems.
13. Quantification of model uncertainty.

In addition to research in these areas, it is becoming increasingly clear that there is much untapped potential in mathematical fields that are not traditionally considered as applied. Good examples are recent applications of algebraic geometry to biological problems and the use of methods from algebraic topology for high dimensional data analysis. (Within SIAM, recognition of these emerging opportunities has led to the establishment of a new SIAM Activity Group in Algebraic Geometry.)

Recommendations – Research Support Mechanisms, Examples of Successful Programs

To support the research areas outlined above, programs at individual agencies and interagency initiatives will be needed. Specifically, **an array of complementary approaches will be needed – from those that focus on building expertise in a single topic area, often at a single agency, to application-driven programs that combine mission agency's user communities and discipline-organized research programs.** Agencies likely to have relevant expertise, communities, programs, and missions include: the National Science Foundation (NSF), the National Institutes of Health (NIH), the Department of Energy (DOE), the U.S. Department of Agriculture (USDA), the Department of Defense (DOD), the Environmental Protection Agency (EPA), the Department of Homeland Security (DHS), and others.

The National Science Foundation has been a leader in the development of models for stimulating and funding interdisciplinary research in general and as it relates to biology in particular. There are several existing programs that effectively support research at the interface of the life sciences on the one hand and mathematics, the computational sciences, and statistics on the other. **These programs could be expanded or used as models for the establishment of new programs at NSF or other agencies**.

One particular inter-agency program has been very successful and enormously valuable to research at the interface of mathematics and biology. **The *Joint DMS/NIGMS Initiative to Support Research in the Area of Mathematical Biology* is a collaborative program between NSF and NIH**, originally established in 2001 and is now in its second 5-year cycle. (A recent meeting of investigators supported by the program over the course of its existence, organized jointly by NSF and NIH, showcased some of the projects that have been funded and demonstrated the truly innovative nature of the program.) **The key characteristic of this program is that it is one of the very few existing programs at any of the Federal funding agencies that allows for new biological AND new mathematical research to be conducted *at the same time within the same project proposal***. (While the program has been very successful, an ongoing concern is that award sizes are too small to tackle larger-scale ambitious projects.)

This dual approach is critically important because, for many of the new technologies being developed to generate biological data (such as next-generation sequencing or in vivo imaging), we still lack the mathematical and statistical tools needed to analyze and interpret these data so that they can be used to increase our understanding of biological systems and provide input for the construction of predictive models. To fully and efficiently tap the expertise of all the different kinds of researchers in this equation—e.g. the mathematicians developing data analysis algorithms, the engineers developing imaging technologies, and the life scientists defining the questions about biological system functioning—**the federal government should be looking for ways to support the development of all elements of a research problem (the tools, models, and experiments) in tandem**. (I will discuss this point more in the section below on the Virginia Bioinformatics Institute and effective environments for interdisciplinary research).

In a related, but broader area, NIH and NSF announced a new program this spring, *New Biomedical Frontiers at the Interface of the Life and Physical Sciences*. While no projects have been selected and funded yet by this new program, the emphasis in the solicitation on supporting efforts that involve multiple investigators who represent the physical, computational or engineering and life or behavioral sciences is to be lauded.

Other examples of exemplary NSF programs include:

- *Cyber-enabled Discovery and Innovation (CDI)* is an NSF-wide initiative established in 2007 and designed to fund projects that use innovation in computational thinking to make advances in any discipline supported by the agency. (At NSF, computational thinking is defined as encompassing computational concepts, methods, models, algorithms, and tools.) This program encourages researchers to think boldly about challenges in data, complexity, and collaboration across multiple disciplines without being constrained by disciplinary cultures and programs.

- *Frontiers in Integrative Biological Research*, a program, phased out in 2008, was designed to support integrated teams of researchers from different scientific fields, focused on biological problems that transcend traditional disciplinary boundaries.

- *Algorithms for Threat Detection*, a joint program between the NSF Division of Mathematical Sciences and the Defense Threat Reduction Agency in DOD, is intended to support the development of the next generation of mathematical and statistical algorithms and methodologies in sensor systems for the detection of chemical and biological materials.

Mechanisms should be available to support a variety of sizes of research projects, from individual investigators to center-scale collaborations. Examples of multi-agency and single-agency center-scale initiatives in this area include:

- The National Institute for Mathematical and Biological Synthesis (NIMBioS), jointly supported by the NSF Biological Sciences Directorate and DMS, together with USDA and DHS.
- NSF DMS supports the Mathematical Biosciences Institute (MBI) at the Ohio State University.

Both institutes focus on research at the interface between the mathematical and computational sciences and biology and foster interactions between mathematical scientists and bioscientists.

Thus, NSF has developed and tested successful models to foster interdisciplinary research at the interface of biology and computation, both within the agency and in collaboration with other Federal funding agencies. These can serve as models for the broader cross agency funding structure advocated by the New Biology report.

In addition to programs that support research activities, federal agencies should focus on raising awareness in the biological and mathematical communities about science at the interface and facilitating cross-disciplinary collaborations, as creating research teams and partnerships across disciplines takes more time and conversation than building teams of people who are within a discipline and share a common culture (this point is discussed in more depth later in my testimony). In addition, outreach within each community about interesting results in one discipline that may potentially be relevant to problems in the other discipline could have a significant impact (i.e. the discovery of applications of algebraic geometry to biological problems mentioned above). Such unexpected linkages can bring very high returns, and their development should be systematically fostered and supported.

To accomplish the above goals, programs that support network creation, workshops, travel, and summer programs, would be useful. "Sabbatical" cross-disciplinary opportunities for researchers, post-doctoral students, and graduate students also might be effective in creating a new community of researchers more alert to and equipped to conduct interdisciplinary research.

An example of a federal effort focused on enabling the creation and sustaining of connections between researchers with common interests is the NSF Research Coordination Networks program, which in 2010 is expanding to include a special track supporting networks of researchers focused on problems at the interface of the biological and mathematical or physical sciences.

INTERDISCIPLINARY COLLABORATIONS – CULTURE AND CROSS-AGENCY COORDINATION

Second Set of Questions from the Committee: What is the nature of the interactions and collaborations between mathematicians and biological scientists at the Virginia Bioinformatics Institute (VBI)? How is VBI facilitating these interdisciplinary collaborations and what lessons can we learn from VBI? Is research at the intersection of the biological sciences, the physical sciences, and engineering being effectively coordinated across the Federal agencies? If not, what changes are needed?

Much of the scientific research in biology and related disciplines happens at universities. By and large, the **nature of the interactions among scientists from different disciplines is constrained by existing academic administrative structures**, which generally do not encourage interdisciplinary research. This has been well documented in the 2004 National Research Council report "Facilitating Interdisciplinary Research,"[5] which also puts forward

5. *Facilitating Interdisciplinary Research* (2004) is available at http://www.nap.edu/catalog.php?record_id=11153.

solutions to this part of the problem. Many universities are addressing the issue of interdisciplinary research by creating research centers that are more flexible administratively and are sometimes organized in a problem-centric rather than discipline-centric way. Some of these centers are "virtual," in the sense that the researchers all have primary appointments in academic departments, with some shared research infrastructure. Other centers have dedicated buildings that provide primary laboratory space. **The institute I work in is part of Virginia Tech's response to the challenge of fostering interdisciplinary research on its campus**.

The Virginia Bioinformatics Institute (VBI) was established on the campus of Virginia Tech in 2000 and is focused on research at the interface of the experimental and computational sciences. The institute currently has a staff of approximately 230, including approximately 150 scientific personnel and a dedicated 130,000 square foot. building, completed in 2004, with in-house computational and data generation cores. Researchers at VBI are engaged in a wide range of interdisciplinary research projects that bring together diverse disciplines such as mathematics, computer science, biology, plant pathology, biochemistry, systems biology, statistics, economics, medicine, and synthetic biology.

My own research is focused on systems biology, in particular the development of mathematical algorithms related to the modeling of molecular networks. My research group has worked on applications to understanding gene regulatory networks, infectious diseases, and, more recently, cancer. During my eight years at VBI I have collaborated with experimental biologists, biochemists, and computer scientists, both at VBI and elsewhere. Based on my experience, **the single most important factor for making VBI an excellent environment for interdisciplinary research is the fact that a wide range of disciplines are brought together under one physical roof**. I am trained as a mathematician and most of my research group consists of mathematicians. But a statistical geneticist occupies the office on one side of me, and my neighbor on the other side is a biochemist. Similarly, my Ph.D. students might share office space with experimental biologists or computer scientists. The two most important benefits of such an arrangement are that, firstly, it becomes very easy to share information. Even in this age of instant electronic access to information and video chats with colleagues around the world nothing can replace a face-to-face conversation or chance encounter at the proverbial water cooler. Secondly, sharing physical space on a daily basis allows for the merging of different scientific cultures. In my opinion, **the most important and difficult challenge in fostering interdisciplinary research is the creation of a common culture and a common language**, even at the most basic level. In a mathematician or a physicist, the word "vector" might elicit the image of an arrow depicting the direction and velocity of a moving object, whereas in a biologist the same word might bring to mind the image of a disease-carrying mosquito or a rat.

A common obstacle in applying quantitative data analysis methods effectively in life sciences research is that biological experiments are often designed without the involvement of a modeler or bioinformatician or statistician. Once the data from these experiments are generated, often at considerable cost, they sometimes turn out to be unsuitable for the desired data analysis or modeling method. It is important, therefore, to assemble the entire team for a project ahead of time, so that everybody can contribute to all phases of the project. The laboratory of one of my collaborators, for instance, is just across the hall from me and I can easily provide input, suggestions, and answers to questions, as I visit frequently. In fact, computational modeling and analysis will become an increasingly important component of the experiments themselves and their design. An integrated environment such as VBI makes the transition to "computer aided design" of experiments easier. It also facilitates biologists' input into the subsequent generation of biological hypotheses through computational methods.

A thorny problem in creating an interdisciplinary environment, one that we have struggled with for a long time, is performance evaluation. In a scientifically more homogeneous academic department it is easier to evaluate the quality of someone's research, since colleagues are more familiar with the different scientific journals in the field and their quality. A common and problematic practice is to replace this domain knowledge with metrics such as the impact factor of a journal. It is well known that it is possible for a journal to influence its impact factor in ways that do not reflect its actual scientific importance. Also, cultural factors in different scientific communities affect this metric. For instance, while *Science* and *Nature,* two of the very best journals in the physical and life sciences, have very high impact factors, the top rated mathematics journals, such as *Annals of Mathematics*, have impact factors that are an order of magnitude smaller. So the impact factor of journals can be only one of several measures to be used. Extramural funding through grants and contracts is another factor that is commonly taken into consideration in academic institutions. **Preparing grant applications for interdisciplinary research tends to take considerably more time and effort than single investigator grants**, and budgets typically need to be larger. Since there are fewer funding programs available for interdisciplinary research than for research within a single discipline, success rates tend to be lower. It is important to provide incentives for scientists to nonetheless embrace interdisciplinary research.

At VBI we are continually working to refine our evaluation process that takes these and other factors into account. For instance, the institute also wants to encourage its scientists to engage in entrepreneurial activities to ensure that their scientific discoveries translate into tangible products that benefit society. So entrepreneurial activity is another criterion in our evaluation process.

The most important lesson I can draw from VBI's experience is that **integration** of different areas of expertise into one physical and administrative structure that is problem centric rather than discipline centric **can serve as an important accelerator of interdisciplinary research**. While this is common practice in industry, it is less so in academe. But it resonates well with the central theme of integration in the New Biology report.

I frequently serve on grant review panels for several agencies, including the NSF, NIH, the postdoctoral program for Federal research laboratories run by the National Academy of Sciences, and a variety of foreign funding agencies. Panels I have served on have focused on a wide range of disciplines, including mathematics, biology, engineering, computer science, oncology, and several interdisciplinary panels. In addition to these agencies, the Office of Science within the Department of Energy, and the U.S. Department of Agriculture also support research at the interface of biology and the computational sciences. In my experience as a reviewer, I have come to realize, that such research takes place in a large variety of settings, including academic departments such as biology, computational biology, biochemistry, physics, bio- and biomedical engineering, electrical engineering, systems engineering, computer science, mathematics, to name the most common ones, as well as a variety of academic and nonacademic research centers, medical schools, government laboratories, and companies. My experience shows me that the scientific community is already mobilizing on a broad scale to meet the challenges outlined in the New Biology report.

While this diversity of computational biology research is a very encouraging sign, it also represents a challenge to funding agencies that need to tailor programs to the different communities. I have described earlier some examples of funding programs that cross disciplines within agencies or span across agencies. The agencies are tapping into a broad and partly overlapping pool of reviewers. It happens to me frequently, that I meet somebody at an NSF review panel, who I had met a few months before at an NIH study section, for instance. And program officers from different funding agencies communicate with each other regularly, in my experience. However, **there are still many opportunities for the agencies to coordinate programs, and a particular need is to pool resources for funding larger-scale projects**. We now have some good case studies we can draw on of programs that create synergy between agencies' expertise, such as the DMS/NIGMS program I mentioned earlier, and can, as discussed in the previous section, be a model for larger-scale cross-agency activities.

Lessons Learned about Interdisciplinary Collaboration and Cross-Agency Coordination

- From our experience at VBI, it is clear to me that integration of different areas of expertise into one physical and administrative structure that is problem centric rather than discipline centric can serve as an important accelerator of interdisciplinary research. The value of co-location is at least two-fold: (1) It allows researchers to develop a common culture and learn each other's language; and (2) It allows multiple disciplines to contribute to the development of hypotheses, the methods for making predictions, and the design of experiments from the beginning of a project.

- One of the major challenges facing interdisciplinary research is that of performance evaluation. One growing problem is how those in a discipline can assess the quality of research of someone publishing outside that field. Another problem is the greater time for preparing proposals to support large interdisciplinary teams and the lower success rate for such large grants.

- Finally, from my experience with multiple federal agencies as a grantee and a reviewer, I am pleased to report that I see good individual collaborations among these agencies – the program officers communicate regularly with each other, the expertise of reviewers are tapped and shared across agencies, and a number of joint programs have been established (as highlighted in the previous section). However, there are still many opportunities for the agencies to coordinate programs, and a particular need is ways to pool agency resources to allow the funding of larger-scale projects.

WORKFORCE - EDUCATION AND TRAINING

Third Set of Questions from the Committee: *What changes, if any, are needed in the education and training of undergraduate and graduate students to enable them to work effectively across the boundaries of the physical sciences, engineering, and the biological sciences without compromising core disciplinary depth and understanding? Specifically, what recommendations* *or changes, if any, would you offer regarding the portfolio of education and training programs supported by NSF?*

As Director of the VBI Education and Outreach Program I devote part of my time to education and training in computational biology from the K-12 to postgraduate levels, in formal and informal settings. The program has four full-time staff members, in addition to myself, including one at the Ph.D. level.

Graduate Education

I will first address education at the graduate level. As the New Biology report states: "Certain institutions have recognized these limitations of traditional departments for establishing the New Biology, and have responded not by eliminating departmental structures, but rather by supplementing or overlaying them with interdisciplinary programs or institutes that have both research and educational objectives. **Virginia Tech** is one of those institutions. In 2003, we **created a Ph.D. program with the name "Genetics, Bioinformatics, and Computational Biology (GBCB)" that was designed to train students at the interface of experiment and computation in the life sciences**. The program is administered by the Graduate School and draws on faculty from several departments and institutes, including VBI. While the program was one of a handful at the time, there are now a number of such Ph.D. programs at other institutions in the U.S. and worldwide. The structure of the program is fairly typical, with each student choosing a major area of expertise, such as computer science or one of the life sciences, together with topics from other minor areas of expertise, and a dissertation research project that involves more than one area. In designing the program, we tried to strike a balance between the need for diversity and depth of training. Other programs may strike this balance in more or less different ways, with varying administrative structures. **Our graduates are sought after in both academic institutions and industry and have no difficulties finding attractive employment opportunities**.

Most of the research in my group is such that it typically requires fairly deep training in mathematics, so that most of my Ph.D. students are enrolled in the mathematics Ph.D. program. (In fact, I have had excellent experiences also with postdoctoral mathematicians with no prior background in biology, who have acquired significant biology skills in a short period of time and have made important research contributions.) In order to learn the requisite biology they take courses designed for the GBCB program and, in effect, their course of study could qualify for the GBCB program as well. Most departmental Ph.D. programs are flexible enough to allow students such a diverse plan of study. So both departmental and interdisciplinary Ph.D. programs can be very effective in training students for New Biology research. **An important prerequisite for the success of departmental programs in this endeavor is, again, integration**. In addition to integration **of curricula**, students need to have an opportunity to **develop a common culture** with other disciplines.

While Virginia Tech has had great success with the GBCB program and other interdisciplinary graduate programs, creating and maintaining such programs is a major investment of time and resources on the part of the institution and its faculty. To date, the NSF Integrative Graduate Education and Research Traineeship Program (IGERT) program has played an important role in creating integrated graduate programs across the scientific spectrum at universities across the

U.S. For example, Virginia Tech currently has four IGERT awards, and their cumulative effect is beginning to transform the institution.

To educate the future scientists who will be critical in realizing the New Biology, universities will have to transform graduate education in many areas, some interdisciplinary, some not. While the IGERT program is excellent at supporting the creation of programs at *newly* established interdisciplinary boundaries, **academic institutions and departments will also have to revisit existing disciplinary programs and established interdisciplinary areas (e.g. the intersection of biology and mathematics). Support from NSF for these efforts—**such as for the design of the structure and curricula associated with such programs, faculty development and training, and the development, coordination, and execution of related activities such as internships, laboratory rotations, fieldwork, and seminars**—would enable universities to create integrated, flexible programs**, as described above, that will prepare the next generation of researchers for the New Biology and other emerging opportunities. The graduate experiences developed by this sort of federal program will benefit multiple disciplines and application areas, and hence such a program may be appropriate for cross-agency partnerships and collaborations.

Undergraduate Education

At the undergraduate level the two most important factors, in my experience, for New Biology training, **are an integrated curriculum and research experiences**. In order **to create an integrated curriculum there is a tremendous need for faculty professional development, especially at the many undergraduate institutions**. For instance, a few weeks ago I lectured at a weeklong workshop for college faculty, entitled “Mathematical Biology: Beyond Calculus,” which was supported by the Mathematical Association of America and was held at Sweet Briar College in Virginia. The participants came from undergraduate teaching institutions around the country, and some came in teams of two: a biologist and a mathematician. The goal was to develop integrated teaching modules that faculty could use in both mathematics and biology classes, and to plan curricula for integrated courses. In my opinion, many more workshops of this type across all the disciplines contributing to the New Biology are needed to allow faculty to develop and teach courses that will interest students in this area and prepare them for interdisciplinary graduate study and research.

Beyond such professional development workshops, **teaching institutions could benefit additionally from close partnerships with research institutions that incorporate professional development, expertise in curriculum development, and research opportunities for faculty and students**. This will enable faculty at these institutions to keep their curriculum up to date, both within and across disciplines, and will allow them to train their students in ways that make them competitive for cutting edge graduate programs. For instance, we are working with three minority-serving undergraduate institutions to set up such partnerships. For the second summer now we are hosting their faculty at VBI where they engage in research and professional development, and we are hosting their students for research experiences. I have found this to be an effective way to help undergraduate institutions keep pace with scientific developments and training needs. It is not clear to me whether there are any funding programs that are particularly targeted at or well-suited to support such partnerships.

The NSF has established the **program *Interdisciplinary Training for Undergraduates in Biological and Mathematical Sciences,*** that addresses curriculum integration and research experiences. The program **is very successful**, in my opinion, **and should be expanded**. It can also serve as a model for similar programs involving other New Biology disciplines. And its scope could be modified to include partnerships of the kind mentioned above.

Genuine research experiences play a tremendously important role in getting undergraduate students interested in the sciences and in preparing them for graduate programs. The NSF's Research Experiences for Undergraduates (REU) program has played an important role in attracting students to science and engineering careers and in preparing them to begin research earlier in their training. For admission to many of the best Ph.D. programs an REU or similar experience has become an important criterion. As I am talking to you here, we have over 30 undergraduates from all over the country at VBI who are doing research with our scientists during the summer, including students from half a dozen states with Representatives on this committee. The students are supported by grants from NSF and NIH. In addition, we have a dozen undergraduates from foreign countries at the institute for the summer. I can see every day what a powerful effect this experience has on the students, and e-mails and letters from past participants make clear that such programs have a lasting impact on them and their career choices.

Recommendations – Graduate and Undergraduate Education

In graduate education, both departmental and interdisciplinary Ph.D. programs can be very effective in preparing students to conduct research in the New Biology, with the key issues being an integration of curricula, the flexibility to strike a balance between the need for diversity and depth of training, and the opportunity to develop a common culture across disciplines. Creating and maintaining graduate programs with these characteristics is a major investment of time and resources on the part of institutions and faculty. Federal support for university efforts to transform graduate education would greatly help prepare the next generation of researchers for the New Biology and other emerging opportunities.

At the undergraduate level the two most important elements for preparing students to work in the areas of the New Biology are an integrated curriculum and research experiences. In order to create an integrated curriculum there is a tremendous need for faculty professional development, especially at the many predominantly undergraduate institutions in the U.S. This could be enabled by programs that support professional development workshops that, for example, bring together faculty from mathematics and biology. In addition, teaching institutions could benefit from close partnerships with research institutions, in which the partnerships provide professional development, expertise in curriculum development, and research opportunities for faculty and students. The NSF programs *Interdisciplinary Training for Undergraduates in Biological and Mathematical Sciences* and *Research Experiences for Undergraduates* have been successful in supporting enhancements in undergraduate education and improving access to critical research experiences, and these programs should be expanded.

Researchers of the Future – K-12 Education and the Perception of Mathematics and Science

Realizing the potential of the New Biology is a long-term effort. It will depend strongly on the generations that are now in the K-12 educational system, their parents who influence their career choices, and their teachers who prepare them for those careers. **There is a tremendous need for teacher training and for providing children with opportunities to experience practitioners of science, engineering, technology, and mathematics (STEM) as what they are: explorers of fascinating mysteries on the most important frontiers of knowledge**. Without changing the image of the STEM disciplines in the minds of the public and our children, we will not succeed in reversing the trend of ever smaller numbers of students choosing STEM careers.

During the last year we hosted over 5000 K-12 students at VBI and we are carrying out programs that involve hundreds of children, their parents, and teachers, in partnership with other organizations, such as Virginia 4H. In my experience, engagement with science and technology at this level can have a huge payoff in the future. Seeing the excitement and genuine interest on the face of a 9-year-old who, in a lecture hall with 400 other children, stands up and asks an insightful question after listening to a scientist talk about nanotechnology convinces me that the number of students electing to study STEM in higher education can be increased, if all stakeholders work together to affect the needed cultural change. There are wonderful examples of such efforts. The U.S. Science Festival later this year will be a signature event for shining the public spotlight on science, and VBI will do its share in our booth to showcase New Biology research. And there are many other smaller events and programs of this type taking place across the country. But given the size of the challenge and the large potential benefit to the U.S. economy and well being, a national effort may be required to affect the needed cultural change. An example of such a larger-scale program is the 2007-2008 "Year of Mathematics," a massive effort by the German mathematical community to help the public experience mathematics. (The program was funded through a public-private partnership with approximately 11 million Euros.)

CONCLUSION

Enabling and exploiting the intersection between the life sciences and the mathematical and information sciences will have great benefits for society, in health, food, energy, and the environment, as noted in the New Biology report. This alone is a reason for the U.S. to explore and invest in this area. However, like in many other fields, such as information technology, medicine, and security, the work in New Biology also has the potential for significant economic benefit to the nation that makes the discoveries and is first to turn them into products and services. The U.S. is not the only nation to see the potential of this area,[6] and the investments of other countries in their research and education infrastructures to produce 21st century innovations lend urgency to our efforts to improve our own research and training capabilities.

---

6. For a discussion of international efforts, see the WTEC *International Assessment of Research and Development in Simulation-Based Engineering and Science*, which includes a chapter on Life Sciences and Medicine, available at http://www.wtec.org/sbes/SBES-GlobalFinalReport_BW.pdf.