Appendix 1

STEERING COMMITTEE

Peter Arzberger (NSF Liaison)
Statistics and Probability
National Science Foundation
1800 G Street, N.W.
Washington, D.C. 20550

Nicholas Cozzarelli
Dept. of Molecular Biology
229 Stanley Hall
University of California, Berkeley
Berkeley, CA 94720

Robert Eisenberg
Dept. of Physiology
Rush Medical College
1750 West Harrison St.
Chicago, IL 60612

Marcus W. Feldman
Dept. of Biological Sciences
Stanford University
Herrin Hall
Stanford, CA 94305

Daniel Kersten
Dept. of Psychology
Elliott Hall
University of Minnesota
75 E. River Rd.
Minneapolis, MN 55455

Simon A. Levin (Chairman)
Section of Ecology & Systematics
347 Corson Hall
Cornell University
Ithaca, NY 14853

Deborah Lockhart (NSF Liaison)
Special Projects Program
National Science Foundation
1800 G Street, N.W.

Washington, D.C. 20550

Andre Manitius
Electrical & Computer Engineering
4400 University Drive
George Mason University
Fairfax, VA 22030

Colleen Martin (editorial assistant)
Ecology & Systematics
347 Corson Hall
Cornell University
Ithaca, NY 14853

Alan S. Perelson
Los Alamos National Laboratory
Mail Stop K710
Los Alamos, NM 87545

Charles Peskin
Courant Inst. of Math. Sciences
251 Mercer St.
New York, NY 10012

Michael Reed
Dept. of Mathematics
136 Physics Bldg.
Duke University
Durham, NC 27706

Wendy Silk
Dept. of Land, Air, and Water Resources
Hoagland Hall
University of California, Davis
Davis, CA 95616

Appendix 2

WORKSHOP ATTENDEES

Joan Aron
Dept. of Population Dynamics

Johns Hopkins School of Hygiene and Public Health
615 N. Wolfe St.
Baltimore, MD 21205

Peter Arzberger (NSF Liaison)
Statistics and Probability
Natinal Science Foundation
1800 G Street, N.W.
Washington, D.C. 20550

Howard Berg
Dept. of Cellular and Developmental Biology
Harvard University
16 Divinity Ave.
Cambridge, MA 02138

Carlos Castillo-Chavez
Dept. of Plant Breeding & Biometry
341 Warren Hall
Cornell University
Ithaca, NY 14853

Joel E. Cohen
Rockefeller University
1230 York Ave., Box 20
New York, NY 10021

Robert Costanza
Chesapeake Biology Lab.
University of Maryland
Box 38
Solomons, MD 20688-0038

Nicholas Cozzarelli
Dept. of Molecular Biology
229 Stanley Hall
University of California, Berkeley
Berkeley, CA 94720

Leah Edelstein-Keshet
Dept. of Mathematics
#121-1984 Mathematics Road
Vancouver, B.C. V6T 1Z2 Canada

Robert Eisenberg
Dept. of Physiology
Rush Medical College
1750 West Harrison St.
Chicago, IL 60612

Bard Ermentrout
Dept. of Mathematics & Statistics
University of Pittsburgh
Pittsburgh, PA 15260

Robert H. Gardner
Oak Ridge National Laboratory
Bldg. 1505 MS 6036
P.O. Box 2008
Oak Ridge, TN 37831-6036

John Guckenheimer
Center for Applied Mathematics
305 Sage Hall
Cornell University
Ithaca, NY 14853

Richard Hudson
Dept. of Ecology and Evolutionary Biology
University of California, Irvine
Irvine, CA 92717

Fern Hunt
Dept. of Mathematics
Howard University
Washington, D.C. 20059

Stuart Kauffman
Dept. of Biochemistry and Biophysics
School of Medicine G3
University of Pennsylvania
Philadelphia, PA 19174

Daniel Kersten
Dept. of Psychology
Elliott Hall
University of Minnesota
75 E. River Rd.
Minneapolis, MN 55455

David Kingsbury
Dept. of Microbiology
George Washington University Medical Center
2300 Eye St., N.W.
Washington, D.C. 20037

Nancy Kopell
Dept. of Mathematics
Boston University
111 Cummington Street

Boston, MA 02215

Eric S. Lander
Dept. of Mathematics
Harvard University
151 Bishop Richard Allen Dr.
Cambridge, MA 02139

Simon A. Levin
Ecology & Systematics
347 Corson Hall
Cornell University
Ithaca, NY 14853

Michael Levitt
Dept. of Cellular Biology
Stanford University Medical Center
Palo Alto, CA 94305

Deborah Lockhart (NSF Liaison)
Special Projects Program
National Science Foundation
1800 G Street, N.W.Washington, D.C. 20550

Don Ludwig
Dept. of Mathematics
121-1984 Mathematics Rd.
University of British Columbia
Vancouver, BC V6T 1Z2 Canada

Marc Mangel
Dept. of Zoology
University of California, Davis
CA 95616

Andre Manitius
Electrical & Computer Engineering Dept.
4400 University Drive
George Mason University
Fairfax, VA 22030

James D. Murray
Centre for Mathematical Biology
Mathematical Institute
University of Oxford
24-29 St. Giles'
Oxford OX1 3LB, U.K.

Garrett Odell
Dept. of Zoology, NJ-15

University of Washington
Seattle, WA 98915

Richard Olshen
Stanford University
School of Medicine
HRP Bldg., Rm. 110A
Stanford, CA 94305

George F. Oster
Division of Entomology
201 Wellman Hall
University of California, Berkeley
Berkeley, CA 94720

Alan S. Perelson
Los Alamos National Laboratory
Mail Stop K710
Los Alamos, NM 87545

Michael Reed
Dept. of Mathematics
136 Physics Bldg.
Duke University
Durham, NC 27706

John Rinzel
National Institutes of Health
Public Health Service
Dept. of Health & Human Services
Bldg. 31, Room 4B-54
Bethesda, MD 20892

Jonathan Roughgarden
Dept. of Biological Sciences
Stanford University
Stanford, CA 94305

David Rumelhart
Dept. of Psychology
Stanford University
Stanford, CA 94305

David Schimel
Natural Resource Ecology Lab.
Colorado State University
Fort Collins, CO 80523

Lee A. Segel
Dept. of Applied Mathematics

The Weizmann Institute of Science
Rehovot 76100 Israel

Wendy Silk
Dept. of Land, Air, and Water Resources
Hoagland Hall
University of California, Davis
Davis, CA 95616

Terry Speed
Dept. of Statistics
University of California, Berkeley
Berkeley, CA 94720

Sylvia S. Spengler
Human Genome Center
459 Donner Laboratory
Lawrence Berkeley Lab.
University of California, Berkeley
Berkeley, CA 94720

Marcy K. Uyenoyama
Dept. of Zoology
Duke University
Durham, NC 27706

Michael S. Waterman
Dept. of Mathematics
University of Southern California
University Park
Los Angeles, CA 90089

James H. White
Dept. of Mathematics
University of California,
 Los Angeles
Los Angeles, CA 90024

Grace Yang
Dept. of Mathematics
University of Maryland
College Park, MD 20742

Appendix 3

Research Opportunities in Computational Biology

Executive Summary

Computational Biology is emerging as a current analog to the development of molecular biology as a discipline in its own right. In the late 1950s and early 1960s a group of scientists began to apply the tools of several disciplines; genetics, microbiology, physics, biochemistry, and biophysics, to analyze biological problems in a new way. The power of this approach was so great that it emerged as a discipline itself and is now known as molecular biology. As described in this document, the application of mathematical and computational tools to all areas of biology is producing equally exciting results, is providing insights into biological problems too complex for traditional analysis, and is emerging as a new discipline within the biological sciences.

There is a consensus among all observers that biology, regardless of the sub-speciality, is overwhelmed with a large amount of very complex data. However, what sets biology apart from other data rich fields is the complexity rather than the sheer volume of the data produced. In contrast to other data rich fields, biology remains a scientific "cottage industry," with the data generation done in a highly distributed mode, with no standard format or syntax.

Thus, all areas of the biological sciences have urgent needs for the organized and accessible storage of biological data. Generally this is referred to as biological database development, however, this terminology infers traditional database technology such as transaction oriented relational database systems. Unfortunately, relational database technology is inadequate to serve many areas of the biological sciences due to the complexity of biological data and the absence of a standardized data structure. It is clear that collaboration between computer scientists and biologists will be necessary to design information platforms which accommodate the needs for variation in the representation of biological data, the distributed nature of the data acquisition system, the variable demands placed on different data sets, and the absence of adequate algorithms for data comparison, which forms the basis of biological science.

There have been dramatic advances in commercially available hardware over the past few years and it has had an effect at both the high and low ends of the spectrum. In the past this general purpose hardware was inadequate to address the most computationally intense problems in the biological sciences. These computational problems were best be handled by special purpose equipment designed by teams of biologists and chip and circuit designers. This condition has been dramatically altered in the past two years as high performance general purpose instruments have become more widely available. Not only hardware limitations have affected the productivity of the computational biologist. There is a continuing need for new algorithm development to cover many tasks, especially comparisons between objects and images. Imaging technology is central to almost all of biology and data representation through image construction remains an elusive but astoundingly powerful tool. The full utilization of modern CAD tools in computational biology will advance image analysis, but will require intense software and hardware development because of the complexity of biological data

During the last decade there were dramatic advances in instrumentation and related methodologies for both light and electron microscopy. The advances lie not simply in higher resolution, but rather in a broader size range of structures that can be analyzed, and more powerful methods for putting together the pieces of three-dimensional puzzles of cell form, and the addition of dynamic details of biological form and function, ranging from the subcellular to the physiological level. The new approaches are computationally demanding. Extant computational resources, which were typically set up for entirely different processing needs, not surprisingly, are proving inadequate for dealing with the massive data flow. An effort to develop new computational approaches is underway in a few laboratories around the world. However, it is important that new software be developed within the context of the experimental research driving the needs; that is, there must be close collaboration between those developing the software and the groups carrying out research on static and dynamic structures. Furthermore, augmentation of the experimental environment, particularly image processing equipment and other specialized equipment, is needed. Positions for sophisticated programmers are even more important. A prime example of the need for such a laboratory-based specialized programming effort is the development of workstations for interactive visualization and interpretation of 3-D data. The development will proceed in pace with experimental research only if it is done in an environment "open" in the terms used by the computer science world, where new applications are developed free from proprietary restraints and distributed as source code to other laboratories facing the same experimental needs. Commercial interests or specialized production groups will be required finally, to add value to the base line development, producing highly reliable ("bullet proof") production line products.

X-ray crystallography and NMR are the major experimental methods for deducing macromolecular structures at atomic resolution. NMR and X-ray crystallography both produce extremely large amounts of data and are entirely dependent upon the availability of powerful computers and sophisticated processing algorithms for the interpretation of raw data. In addition, there are fundamental scientific problems in both areas that require major computational advances. In addition, substantial opportunities exist for combining structural information from several experimental techniques. This may provide the basis for a structural solution where only partial data are available from any single technique.With improved computational tools, combining physical data from a variety of sources may become common place. These developments will allow solutions to be obtained for structural problems which would otherwise be intractable. Analysis of errors in structures based upon experimental data from several sources also represents a new computational challenge.

Advances in X-ray and NMR data analysis will lead directly to rapid developments in the field of protein folding which will be synergistic with developments in other areas of biology itself, and especially computational biology. Common problems of data representation, search strategy, pattern recognition and data visualization appear in many fields. There is a particularly exciting synergistic relationship between the protein folding field and those of structure determination by X-ray crystallography and 2-D NMR. Each field will benefit from rapid advances in the other disciplines. Improved folding algorithms provide a new way to attack the phase problem in crystallography, and new, more carefully refined protein structures provide rich new insights into protein folding.

Various initiatives in computational neurobiology give us the hope of interpreting the mass of anatomical and physiological information about the nervous system that is now available in functional terms. Better interpretation of these data will permit neurobiology to make contact with other fields such as psychology and artificial intelligence. This work will make specific, testable predictions in the areas of sensory perception (visual, olfactory, and auditory), memory, learning, and motor control. Above all, it will lead to the integration of all these aspects to provide an eventual understanding of the total functioning of the nervous system. Such integration can be expected to provide new insights that will lead to improvements in the treatment of diseases of the nervous system at all levels, from neuropharmacology to psychotherapy. In addition, studies of this kind may be expected to contribute to major advances in artificial intelligence and practical robotics.

In the area of genome analysis significant progress has been made over the past few years, including the use of molecular tools such as Restriction Fragment Length Polymorphism (RFLP) analysis. However, considerable effort is still required to make genetic linkage maps effective tools for genetic research. To be useful in common situations, more markers must be identified and mapped to produce higher-resolution maps. In many cases marker analysis requires the ability to analyze small families and consider quantitative traits. To be fully useful in a meaningful quantitative sense this analysis will require powerful computer simulation and modeling. Common to all of the problem areas examined is the need for good visualization of data. Visualization is necessary because the sequence analysis phase for a molecular biologist is equivalent to exploratory analysis for a statistician. It is at this point that the experimentalist gains the feeling for, and understanding of, a sequence which may then guide many months of experimental work. The complexity inherent in biological systems is so great that very sophisticated methods of analysis are required. These are the tools which must be readily accessible to molecular and cellular biologists untrained in computer technology.

Ecology and evolutionary biology encompass a broad range of levels of biological organization, from the organism through the population to communities and whole ecosystems. This complexity demands computational solutions. The need for enhanced computational ability is most evident when one attempts to couple large numbers of individual units into highly interactive and largely parallel networks, whether at the tissue, community or ecosystem level of organization. The proliferation of information from remote sensing introduces the need for geographical information systems that provide a framework for classifying information, spatial statistics for analyzing patterns, and dynamic simulation models that allow the integration of information across multiple spatial, temporal, and organizational scales. Today, in these fields application software is mostly nonexistent except in a few special special cases such as image processing and remote sensing. As more researchers begin to use computational techniques, we can expect to see a wider sharing of applications developed by an individual or small group. This will require additional resources to take research codes and make them "bullet-proof" enough for community use and to add adequate documentation. In order to take advantage of all these new capabilities, we need to increase training modalities. This can take a wide variety of forms, from on-line self training techniques to special sessions at universities, national centers, or workshops.

It is recommended that Federal granting agencies place greater emphasis on the area of Computational Biology through a number of mechanisms. This support must be developed over a period of several years with a particular emphasis on infrastructure and training. Many of the necessary changes may be instituted immediately while others will require a longer time in order to generate budgetary resources to build in new areas. The current focus on biological databases is a good beginning, however, the need is so great that the initiative needs considerable additional resources. These resources should be directed in three areas. First, the enhancement of current databases which are in wide use but need concerted effort at standardization of data structures and broadened access. Second, a continued examination of new databases which will incorporate important information needed by many investigators, but also explore new database ideas and representations. Third, research on the representation of objects and images which will be searchable and comparable within database structures. For example, there is a great need to be able to search a database of enzyme or antibody active site configurations to test for binding of newly developed ligands. Database development remains the highest priority item since this area is common to all fields of biology. A second area of high priority is the development of more powerful visualization tools for data interpretation. This area too is a need shared by almost all fields of biology. Funding agencies could immediately respond to some of the needs of the research community by recognizing the need for professional programmers and hardware and software facilities on grants in this area. Agencies must break out of the habit of immediately removing these items from budget requests in order to reduce the overall cost of an award, since these items are critical not only to doing the proposed work but also to making the results of the work (in the form of usable source code) available to the rest of the research community.

Appendix 4

Training Computational and Mathematical Biologists

Summary

1. Introduction

It has been estimated that in mid-1990, there were approximately 4000 professional level scientists identifiable as computational or mathematical biologists. These scientists were found in a wide variety of institutions and in a wide range of positions within those institutions.

The pattern of distribution of these individuals among and within different institutions appears to be related to their academic training. For example, mathematicians and computer scientists who have primarily followed an interest in the biological sciences generally work as biologists and find themselves in nonacademic research positions in industry, government or private research institutes, or quasi-academic research centers (e.g., supercomputer centers). A small minority are in biology departments. In contrast, mathematicians who have continued to pursue research activities in mathematics, choosing biologically related problems or examples, or collaborating with biologists, tend to remain in departments of mathematics or applied mathematics in academic institutions. Computer scientists follow a similar pattern. Statisticians may be found in statistics departments, biostatistics groups or departments, or even in biological sciences departments, depending on the extent of their involvement with biological problems, and the local structure of the institutions within which they work.

Biologists who rely on computational and mathematical tools in their research activities are found in many institutions. A large number have moved into industry where they play a role in the analysis of macro-molecules in biotechnology and pharmaceutical companies. Another major source of employment is in government and private research institutes, which tend to focus on problem-oriented research and directly utilize their computational biology skills. In the academic environment, computational biologists pursuing accepted biological problems are found in a variety of departments of biology (including departments of related name such as genetics, ecology and evolutionary biology, molecular biology, and microbiology), chemistry, and biochemistry.

The character of the institutional acceptance of these interdisciplinary activities depends on two factors: the need of the institution for problem-oriented work, and the traditional academic expectations on the performance of the individual. For example, biology departments place their emphasis on disciplinary achievements, and computational and mathematical approaches are secondary to the disciplinary results. Therefore, the infusion of mathematical and computational tools is dependent on the confidence of the researcher that they can afford to invest the time and effort to enable them to use this approach, let alone develop new tools. Thus in many cases, computational and mathematical biology makes a backdoor entrance into the academic world. In contrast, these approaches are embraced more directly by industry and research institutes whose problem-oriented programs utilize a broader range of approaches, including direct application of mathematical and computational techniques.

The workshop participants' assessment is that in the immediate future, this situation will not undergo a substantial change. Therefore, scientists expecting to enter the academic research world will continue to need a strong disciplinary grounding for their cross disciplinary work. Employment opportunities in industry and research institutes appear to be stable, or growing slowly. Such centers will continue to be major sites for the development of computational techniques and applications in biology.

Because of their frequently strong mathematical and computational environments, and the less frequent presence of rigid departmental structures, one possible source of future growth for computational biology is the four-year college. Mathematical and computational approaches fit well within the research environments found in these institutions, and they are likely to find effective implementation in the teaching programs. In this context, faculty in these institutions may be expected to employ mathematical and computational techniques in both research and the development of teaching aids that will eventually find their way into research institutions. However, here again, strong disciplinary training will be essential as the basis for the research approach.

2. Profiles of Computational and Mathematical Biologists

In the past, most of the migration of scientists into computational biology has been from disciplines outside of biology (e.g., math, physics, chemistry, computer science, etc.). Physicists become biologists, but not the reverse. This migration and its asymmetry has been prompted by successful applications of domain-specific technology to solving biological problems.

Many early successes in computational biology were obtained by scientists who were primarily biologists with marginal skills in computer science and mathematics (programming skills and some algorithmics), while many others were the result of work by scientists with extensive mathematical and computational backgrounds. However, as the problems under investigation become more complex, training which provides great depth in quantitative analysis will be essential.

Current interest and excitement in computational and mathematical biology is driven in large part by neurobiology, global change, and genomics. In all of these areas, vast amounts of information are accumulating at a rate that precludes human absorption and, hence, understanding. Biology needs tools for manipulating and analyzing information. In order for training environments to be maximally effective there must be a clear understanding of which professional profiles are suitable for current and future researchers in computational and mathematical biology.

The profiles which follow are dependent upon the nature of the position. Academicians tend to reside within traditional departmental units; whereas, in industrial settings and research institutes there is a wider range in the mixtures of disciplines in working groups. The following lists of specialities within computer science, mathematics and biology are those in which there is substantial research activity today and where there is likely to remain some research focus in the future.

Computer Scientists:

Most computer scientists retain their primary professional identification with computer science. They tend to view biological applications as a source of computer science problems. Biological applications are new to computer scientists, and the traditions across the interface are developing at a moderate pace. The tendency is to cross the line as a senior scientist by developing collaborations. There are some successful scientists in this field whose first exposure to biology was at the graduate level. Examples of the areas of computer science in which such collaborations take place are:

Artificial Neural Networks (AI)
Algorithmics
Database design and theory
Visualization (Graphics)

Biologists:

Biologists working on computational problems come from a plethora of backgrounds: computer science, mathematics, statistics, engineering, physics and chemistry as well as biological disciplines. The biological sciences are themselves diverse and different areas of biology draw upon very different quantitative skills. Those biologists who have crossed the boundaries between biology and other disciplines have often done so to address specific biological problems. Their acceptance by the biological community has been out of necessity since many biological problems require technology that has been driven by insight and intuition from other disciplines. This report is motivated by the assumption that this trend will accelerate in the near future in areas such genomics, neurobiology, imaging, structural biology and issues of global climate change. Many of these developments have been initiated by scientists whose initial training was outside biology (e.g., mathematics, chemistry and physics). The current technological advances will require a new range of quantitative skills beyond the norm of current curricula in the biological sciences. Biological Sciences that currently draw substantially from the computational and mathematical sciences include:

Population Biology, including Ecology and Genetics
Molecular Biology
Molecular Genetics
Cellular Biology
Neurobiology
Biophysics and Structural Biology
Ecosystem Ecology
Epidemiology
Physiology

Mathematicians:

There is a long tradition of mathematicians and statisticians working on biological problems. Indeed, the field of statistics grew largely out of biological origins, and there is a substantial portion of the statistics community working on problems of biometry and biostatistics. There is also a small but stable community of mathematical biologists working within departments of pure and applied mathematics. Some members of this community migrate to biological departments during the course of their careers while others remain in mathematical science departments. Those who do remain within mathematical science departments either establish a career based upon collaborations with biologists, or focus upon mathematical questions driven by biological problems. In some cases, threads of mathematical research initiated by biological problems take on a life of their own as interesting areas of mathematics per se. Areas of mathematics making substantial contributions to biology include:

- Applied Mathematics (Differential Equation Models, Image Processing and Analysis)
- Probability (Sequence Analysis, Interacting Particle Systems)
- Statistics
- Discrete Mathematics
- Topology and Differential Geometry

2.1. Summary of the Current Status

With regard to the current panorama of activity, we perceive that several difficulties exist. First, computer scientists are not sufficiently involved in computational biology. Their work is frequently on problems so abstracted from the application as to make them less than fully effective as collaborators. Another limitation is that biologists tend to view the work of computational scientists as service, and not original research, which tends to alienate this community. Mathematicians are caught between mathematical peers who evaluate their work on the basis of its mathematical depth and elegance, and biologists who have little appreciation for theory that does not have a direct bearing on the interpretation of experimental data. Finally, those biologists who have invested in cross-training are frequently misunderstood and undervalued by their colleagues, most of whom do not understand how to evaluate their work.

Computer science is a new discipline that is rapidly maturing. As the field develops, a tradition of interdisciplinary work will evolve much as it has for mathematics, especially statistics. This will, in part, alleviate the problem of computer scientists' involvement. A greater emphasis on the early grounding in scientific disciplines while at the undergraduate level should also help to cultivate computer scientists with a stronger interdisciplinary focus. As the needs for computation in the various areas described above becomes clearer, the biological community must become increasingly more tolerant and accepting of computational biologists within their midst. As a result of this and other factors, such as heavy dependence on physical measurement, the training of biologists at all levels must become increasingly more quantitative in nature.

3. Encouraging Interactions

The most effective way to encourage interactions between mathematicians and computer scientists on the one hand, and biologists on the other, is through direct co-involvement with a particular problem. This applies at all levels from undergraduate through senior scientist. The ways in which this interaction may be encouraged depend on the level and direction of movement (math/CS to biol or biol to math/cs). At present, the pattern is generally unidirectional, with movement from mathematics or computer science into biology as the dominant paradigm. Significant changes in this state of affairs are likely to require substantial curricular changes based upon effective means of overcoming the apprehension of most biology students towards mathematics.

Interaction can be improved through a strengthening of mechanisms that already exist. However, one area deserves much greater emphasis than is now the case, and that is support of small research groups with a genuine interdisciplinary focus: within this, substantial support is needed for post-doctoral scientists. Support of small group research will develop critical mass in important areas, will help to foster and sustain collaborative research, and provide a crucial home for individuals who are in the early stages of (what is now) a cross-disciplinary research career.

The most effective mechanisms for stimulating these fields vary by the level of a scientist's career stage as outlined below.

(a) Senior researchers (tenured and above)

(CS, Math -> Biol) Support for sabbaticals and, later, research in biology.

(Biol -> Math/CS) Support for visits to math research groups to learn/update new technical areas.

(b) Pre-tenure

Most mathematics and statistics PhD students will start in untenured positions. Changing fields (or, at least becoming more interdisciplinary) at such an early stage is a very risky career move, particularly by individuals approaching a tenure decision. One way to ameliorate this situation is through a new focus on PYI-level type support (National Science Foundation Presidential Young Investigator) for promising people (prestigious competitive awards).

(c) Postdoctoral

Support for postdoctoral training within existing grants is essential. Postdocs are an important educational component of existing research groups, and are very scientifically profitable in the short term. These should support a given individual for multiple years, and not be specifically tied to a particular investigator within the group. This mechanism allows quick response to changing areas of interest, while providing enough time for a postdoctoral fellow to develop a useful independent research focus.

Another aid to young investigators is the computational research associates program at the NSF sponsored Supercomputing Centers. This program is of great value to the biological sciences and the field would benefit from its continued existence. However, to be maximally effective these investigators must be part of an active and focused research program and not "generalists" in applied computer science.

The concepts behind these training programs are not based on the assumption that all people passing through them will eventually obtain tenure track positions in universities.

(d) Graduate students

An important source of mathematical biologists comes from mathematically trained undergraduates who change fields early in their postgraduate education. Such students are then main-stream biologists, with the requisite quantitative background to enter the fields of mathematical or computational biology. The educational challenge for students with this background is the continuation of the quantitative approach to biology in a supportive environment. This requires an appropriate mentor and an appropriate departmental or graduate group environment so that the student’s background is valued and prior training reinforced. Given the many opportunities available to an undergraduate with computer science or mathematical training, it is essential that graduate student support be provided to entice these students to forego the immediate gratification of lucrative employment for the longer term prospects of graduate training and research careers in biology. To this end the continued and renewed support of training grants or traineeships (for example in the research groups described above) are of central and continuing importance.

Furthermore, educational institutions must be encouraged to recognize the need for training students in these areas as a means of dealing with the future of biological research. To this end institutional and departmental support of fellowships and RA (Research Assistant) positions are of supreme significance. Cross-training students at the graduate level will lengthen an educational process that already can be inordinately long. Freeing a student from the demands of a teaching assistantship or a research assistantship with responsibilities to further the work of a principal investigator will help make such programs educationally feasible. It would be especially appealing to find a mechanism to support mathematical or computational biologists within the structure of departments of mathematics or computer science.

One of the most significant factors in the training of graduate students is the role model of the major professor. This mentorship plays a greater role in the ultimate aspirations of a student than is generally acknowledged. The successes, failures, and frustrations of a student’s mentor plays a profound role in the expectations and aspirations of a student. In this context the small group research environment is a highly significant environment in which to train students for the future of the biological sciences.

(e) Undergraduate

In most institutions it is very common for the top biology students, especially those interested in eventual graduate study, to participate in undergraduate research projects, especially in their Junior and Senior years. this opportunity should not be confined to biology students, but should be expanded wherever possible to include interested students from mathematics and computer sciences whenever possible. The proper environment is essential to the nurturing of a student that might wish to commit to a career in the biological sciences, using this valuable undergraduate training. To this end the National Science Foundation REU (Research Experiences for Undergraduates) program provides an extraordinary opportunity in the Math/Biol area.

One area of extreme importance for the future development of a cadre of computational and mathematical biologists, and for the continued recruitment of students into biophysics and related disciplines is the development of better course materials devoted to the quantitative approach to biology. The workshop participants valued very highly the concept of "enculturation of quantitative thought" through the introduction of quantitative approaches in biology courses

(f) Pre-college

While there was considerable discussion during the workshop regarding the state of pre-college science education, no specific recommendations were developed. Many private and government agencies have focused great attention on this problem, and it remains a top national priority. There was general agreement that two issues posed particular concern to the participants. First, the need to involve more fully parents in the educational process. This is particularly important in groups which do not have a cultural history of educational achievement. The second concern was the current selection of the "ultimate underachiever" as the folk hero of the nation’s children. We believe that this message is alarmingly inappropriate in the current context of rapid technological change and global competition. The participants hope that the leadership of the Education and Human Resources Directorate of the National Sciences Foundation will use its influence and insight to find a mechanism to reverse this trend.

(g) Summary Principles

(1) If time is limited for education, spend it in mathematics, not computer science.
(2) What we want is an attitude/consciousness change, so that people are aware of the input of the "other" type of science in their own area.
(3) While collaboration will enhance the science of the current generation, we are seeking to change the way that biology is done by changing the way biologists are educated for the next forty years.

4. Fundamental Educational Principles

(a) Undergraduate Education

(i) General Course Content

The cross-disciplinary aspects of modern science must be emphasized in all undergraduate science and mathematics courses. The role of computer science and mathematics, as well as technologies from physics and chemistry, need to be presented in biology courses. In contrast, the research areas that have used various tools of computer science and mathematics in the experimental sciences should be identified throughout mathematics and cs courses.

(ii) Mathematics/Computer Science Majors

All mathematics and computer science majors should have required experimental science courses. We recommend a minimum of two years that can be concentrated in one area or spread over the basic sciences. The purpose of this is to provide the student with an understanding of the vocabulary and concepts and an experience of the ways in which mathematics or computer science have contributed to other disciplines.

(iii) Biological Sciences

In order to produce biological scientists who will be qualified to do modern research, we strongly recommend that the science curricula require four years of mathematics and/or computer science. Representative courses might include programming, theory of algorithms, probability and statistics, linear algebra, calculus, discrete mathematics, and numerical analysis.

(b) Consequences

Failure to implement these recommendations at a minimal level will foreclose the future for many undergraduates majoring in biological sciences. This originates in the types of problems that are coming into existence and that are consistently more and more dependent on quantitative skills for their solution. Secondarily, lack of training in these quantitative areas will limit the questions that can be asked by an investigator, and may come to threaten an individual's levels of funding. We must remember that we are addressing the education of persons who will be in the pool for the next forty years. If education changes are not implemented, much of biology will fail to thrive.

The broad education that we are proposing also permits people to change their minds and acquire additional course work in another field, even late in their studies, without having to start from the beginning.

Our recommendations should not be construed to support any concept that presupposes a gender-specific bias in the ability to perform. It may be that a type of math/cs anxiety will become apparent if our recommendations are instigated. In order to counter this, we propose that support groups, personal tutorials, study circles and other tools of encouragement and enhanced performance/esteem be supported so that they are readily available.

5. Additional Recommendations

Part of the difficulty in implementing the course recommendations may be the prevalence of pre-med education as a major component of biology curricula. Although there will be a number of additional consequences, it would be well worth considering the restructuring of the undergraduate major so that pre-meds follow a separate track and their presence does not determine the future of an academic discipline.

It is incumbent upon those who practice cross-disciplinary science and mathematics/CS to become both role models and mentors for others. It is particularly important for representatives of under-represented groups to make an effort to encourage others.

Several members of the group have suggested that a new type of biology course should be developed. It would cover the elements of modern biology, but highlighting the contributions of other disciplines. The hope is that someone will be inspired to write a founding text, one that will change the field.

Graduate Education

Continue to create opportunities for cross-disciplinary work. NIH programs in molecular biophysics and the NSF research training groups are examples of attempts to encourage this type of interaction.

One-on-one mentor/student relationships are not sufficient to maintain cross-disciplinary development. Direct support for cross-disciplinary efforts would help to break down the interdepartmental barriers that frequently exist. Seminar groups or other frequent interactions should be encouraged.

New graduate students (and postdocs) might acquire an elementary grounding in a new field through summer institutes or some other "crash course." The courses would be taught by highly interactive, expert, senior level researchers. For example, a course in basic molecular biological concepts could include molecular biology, biochemistry, and molecular biophysics. Emphasis would be on the vocabulary and point of view, that is, how the science is done and what are its assumptions. For a course on computation in genetics, this material might include basic computer science concepts, e.g., files, databases, algorithms and their use, graphics and statistics. The benefits of such a course could also be made available to more senior investigators.

6. Women And Other Under-Represented Groups

In high school, women represent a reasonable proportion, approximately 30-40%, of those students who are interested in the physical sciences and mathematics. Partitioning begins in college and is nearly finished by graduate school. Some disciplines within the biological sciences do have equivalent or even over-balanced representation by women. Increasing the level of course work in mathematics and computer science may be threatening to some of these women. In order to prevent this, specific actions may well be necessary. Similarly, for some students from other under-represented groups, it may be necessary to have additional courses available at the undergraduate level to improve the level of computational competence of entering students.

Appendix 4

Acknowledgement

We wish to thank the steering committee for helping to make the meeting a success. The plans and format were developed together with this committee, and were caried out with the assistance of Colleen Martin at Cornell University.

In addition to those in attendance three individuals contributed significantly to the development and refinement of the Workshop Report. We would like to thank Deborah Lockhart, DeLill Nasser, and Judith Sunley of NSF for the time, attention, and constructive additions they made to the report.

The workshop was supported by a grant from the National Science Foundation (DMS 89-10353) to Cornell Unlverslty (Simon Levin, Principal Investigator)

.