The Role of Theory in Advancing 21st Century Biology: Catalyzing Transformative Research

Committee on Defining and Advancing the Conceptual Basis of Biological Sciences in the 21st Century, National Research Council

ISBN: 0-309-11250-8, 208 pages, 6 x 9, (2008)

THE ROLE OF THEORY IN ADVANCING 21ST-CENTURY BIOLOGY

Catalyzing Transformative Research

Report of the Committee on Defining and Advancing the Conceptual Basis of Biological Sciences in the 21st Century

Board on Life Sciences

Division on Earth and Life Studies

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

Contents

Summary

From microorganisms to whales, from single cells to complex organisms, from plants to animals to fungi, from body plans to behavior, the diversity of life is amazing. Living organisms have a profound impact on our physical world of ocean, landscape, and climate; around us is a multitude of diverse ecosystems that provide a livable environment and many valuable resources. The study of life—biology—is a multifaceted endeavor that uses observation, exploration, and experiments to gather information and test hypotheses about topics ranging from climate change to stem cells. The field of biology is so diverse that it can sometimes be hard for one individual to keep its breadth in mind while contemplating a particular question.

This study was initiated at the request of, and with the sponsorship of, the National Science Foundation. It was conceived as a new approach to a question that has been asked before: What is the future of biology? In 1989 the National Research Council released a report on this topic entitled *Opportunities in Biology*. Over 400 pages long and four years in the making, the report provided a detailed snapshot of the state of biology at that time. Eleven different panels detailed the opportunities awaiting the rapidly diversifying field of biology. Reading the report today, the excitement of that time is palpable. Section after section describes new technologies and promises new discoveries. Each section focuses on a different subdiscipline of biology.

This report takes a different approach by looking for commonalities across subdisciplines. The committee was charged with examining the role of concepts and theories in biology, including how that role might differ across various subdisciplines. One facet of that examination was to consider the role of the concepts and theories in driving scientific advances and to make recommendations about the best way to encourage creative, dynamic, and innovative research in biology. The charge was to focus on basic biology, not on biomedical applications.

At the first committee meeting, to begin identifying the theoretical foundations of biology, each committee member discussed the theories and concepts underlying his or her particular area of research and addressed how those theories and concepts might connect across the field of biology. The talks demonstrated that biologists from all subdisciplines base their work on rich theoretical foundations, albeit of very diverse kinds. They highlighted the varied extent to which theories are an explicit focus of attention and discussion. For example, cell theory underpins much research, but the theory itself is rarely the topic of explicit attention in the research literature.

The committee concluded that a more explicit focus on theory and a concerted attempt to look for cross-cutting issues would likely help stimulate future advances in biology. To illustrate this point, the committee chose seven questions to examine in detail. The list of questions is not comprehensive but rather illustrative. The questions, as shown below, were chosen to show that a focus on theory could play a role in helping to address many different types of interesting and important questions at many different levels.

1. Are there still new life forms to be discovered?

New organisms continue to be discovered, some in environments that were once thought incompatible with life. How many new life forms remain to be discovered? What additional strategies for movement, sensation, and chemical synthesis will be found? How diverse are the variations on the patterns of development of organisms' body plans? How do complicated communities of different organisms affect each other's evolution and what can be learned from the diversity of social organizations that have evolved in different species? How is diversity encouraged and limited by environment? For billions of years, life was exclusively microbial—to what degree can a better understanding of that early evolution change our understanding of the present microbial world, which is turning out to be vastly more diverse than ever imagined, and the processes that underlie all life forms?

The diversity of life presents a huge challenge to biologists but also a virtually limitless opportunity. Both the unity and the diversity of life are explained by the theory of evolution: All life forms share many characteristics because all are descended from a common ancestor and life has become diverse through billions of years of descent with modification. However, the extent and implications of all that diversity are not yet fully understood. An enormous amount of productive research has demonstrated many mechanisms by which evolution leads to diversity. However, much remains to be described and explained. There is need for further theoretical insight into how diversity is generated and maintained, not to mention understanding the implications of losses of diversity. These are exciting challenges.

2. What role does life play in the metabolism of planet Earth?

Diverse as life is, the metabolic pathways that support it are, perhaps surprisingly, quite well conserved and are based on just a few basic strategies. These metabolic pathways, which are the means by which organisms acquire the energy and material components they need to survive and reproduce, have a profound global impact as living organisms form part of global geochemical cycles. The Earth today has been shaped in many ways by metabolic processes, which are key molecular processes at the cellular level as well. Understanding the evolution of these pathways, how they integrate, and how living systems are coupled to environmental conditions is a profoundly important question to several areas of biology and on many scales of time and space.

3. How do cells really work?

The living cell is a marvel, containing thousands of interlocked chemical reactions that harvest energy from the environment, synthesize thousands of different chemicals, manage waste, and recycle components. Ultimately, the cell makes a copy of itself. No human factory can rival the cell's compact and coordinated productivity. Only a fraction of its pathways can be reproduced in the test tube. The laws of physics and chemistry apply, of course, to all living organisms. However, most life processes are maintained far from chemical and thermodynamic equilibrium. Thus, understanding how chemical reactions take place in the crowded and highly organized molecular environment of the cell, or how physical variables like temperature and concentration gradients affect and are affected by living processes (for example, during development, or in the cell cycle or circadian cycle, when the instructions encoded by DNA are manifested in physical processes), is a major challenge of biological research. The interfaces between some current research areas of physics, chemistry, and biology that elucidate these questions are expected to be very fruitful.

4. What are the engineering principles of life?

DNA is made up of nucleotides, proteins of amino acids. Organisms contain many types of cells, ecosystems many different species. The hierarchical organization of building blocks at different scales is a common theme in biology, whose evolution is not fully understood. Complicated systems at every scale are made up of simpler modules that vary in definable ways and combine in ways that result in structures capable of much more than the individual parts. This characteristic of complicated structures, functions, and behaviors arising from the combination of simple parts represents an almost universal theme in biology. Furthermore, across all scales of biology, from subcellular circuits to ecosystems, many biological systems demonstrate "robustness": in other words, they continue to function despite defective parts or changes in the environment. Like the workings of the living cell, this robustness is a biological phenomenon that has evolved through variation and selection and that human engineers would be proud to duplicate. Understanding the principles by which modules combine to create systems with particular properties (another useful, cross-cutting concept) will undoubtedly result in theoretical insights that would apply across biological scales from the molecular to the ecosystem—and perhaps provide valuable lessons for human efforts in design and engineering.

5. What is the information that defines and sustains life?

The power of the computer rests in its ability to represent an immense range of phenomena in digital form that can then be manipulated. Many of the characteristics of life can similarly be represented as flows of information, as it is striking that all living organisms and communities of organisms are able to sense, process, remember, and respond to many different kinds of external and internal stimuli that can be conceptualized as information. Evolution, for example, can be viewed as a process whereby selection of variant genomes is affected by the information provided by the environment. In this view, the information defining the relevant environmental variables is partly encoded in the genome of the adapted organisms by the process of selection, and evolution is thus a process of selective memory in molecular form stored in the genomes of living organisms. The complexity of biological systems can be described using the ideas of information science, but there are deeper conceptual problems in making full use of those concepts of information that were developed for engineering and physics, where they are used in pattern recognition, communications, and thermodynamics. In biological systems, information is intimately dependent on context, making it difficult to apply the concept of information in ways that do not oversimplify complicated biological phenomena. Thus, further development of both concepts and tools will be required to realize the potential of this powerful conceptual point of view.

6. What determines how organisms behave in their worlds?

Organisms as diverse as bacteria and humans possess the ability to respond to their environments and to shape their behaviors in response to specific environmental variables. Understanding how organisms live requires determining the rules that govern how organisms behave in their world, how they sense their environments, and how they use this information to change their behavior. It is important to remember that organisms do not simply wait passively for information from their environments. Their physiology is internally generated, by genetically determined rules, and input from the environment is used to alter the behavior of the organism. In addition, much behavior is generated to actively explore the environment in search of specific sensory signals. For example, bacteria have receptor proteins that allow them to sense concentrations of chemicals in their environment and use these gradients to govern their movements. The integration of sensory information into a form that can be processed by the organism, the nature of the processing machinery, the influence of the internal states of the organism, the influence of the experience on the future states of the organism, memory mechanisms, and many other issues have direct relevance to many different biological regimes, scales, and kinds of organisms. There is a remarkable potential for finding commonalities amid the diversity addressed by this question.

7. How much can we tell about the past—and predict about the future—by studying life on Earth today?

The ability of living systems to pass on the directions for reproducing themselves and for surviving in the environments where those offspring will find themselves is fundamental to the living state, and it is more than a loose metaphor to say that organisms' genomes represent an imprint of past environmental conditions, history, and the selection pressures on the ancestors of organisms. The sequences of the genome are not the only records of past conditions; the ways in which those sequences are put to use are also affected by other past conditions that are carried forward by living systems—from stable physiological states, to imprinted DNA that modifies gene expression, to memories stored in the brain and nervous system, and behaviors remembered and taught to descendants. New mechanisms and new applications of this common ability of living things to record information about the past in some physical, molecular form continue to emerge. Thus, the commonality and diversity in the ways in which organisms represent and use this kind of information are very promising and very challenging frontiers for future research. The record of the past is imprinted in both the fossil record and the DNA of today's living world. Whatever life on Earth looks like 1 million years from now, it will evolve from what is currently alive. If scientists truly understood how current organisms and environments interact to produce future generations, could the course of evolution be predicted?

FINDINGS AND RECOMMENDATIONS

Of course, it is impossible to cover all of biology in so short a report. If the average freshman biology textbook needs hundreds of pages to cover the basics, a mere seven questions cannot possibly introduce even a fraction of the exciting and innovative biology research that is currently underway. The questions are meant to be illustrative, not all-inclusive, and should be read not as a guide to the most important or promising areas for future emphasis but as several examples of the way that concepts and theories can connect the different areas of biology. After exploring this set of seven questions, the committee came to consensus on several findings and recommendations that flow from the idea of looking at cross-cutting issues in biology with an eye to the role of theory.

Finding 1

Biological science can contribute to solving societal problems and to economic competitiveness. Basic and applied research targeted toward a particular mission is one way to accomplish this important goal. However, increased investment in the development of biology's fundamental theoretical and conceptual basis is another way to reap practical benefits from basic biological research. Theory is an integral part of all biological research, but its role is rarely explicitly recognized.

The living world presents a vast reservoir of biological solutions to many practical challenges, and biological systems can inspire innovation in many fields. The many ways that basic biological research contributes to medicine are very familiar, but basic biology can also contribute to advances in fields as diverse as food, fishery, and forest production; pest management; resource management; conservation; transportation; information processing; materials science; and engineering. Biological research breakthroughs, therefore, have the potential to contribute to the solution of many pressing problems, including global warming, pollution, loss of biodiversity, fossil fuel dependence, and emerging infectious diseases.

As the many examples in this report attest, biology is characterized by unity and diversity. There is unity because many biological processes have been preserved through evolution. There is also diversity because natural selection has led to many innovative solutions to the practical problems that living organisms have encountered over billions of years. Therefore, discoveries about a particular organism, sensory pathway, or regulatory network can have immediate applications throughout biology, and the transformative insight that provides the most direct path to a practical solution may arise in a seemingly unrelated research area. Giving explicit recognition to the role of theory in the practice of biology and increasing support for the theoretical component of biology research are ways to help make such connections and thus leverage the value of basic biological research.

The extent of life's diversity has not yet been plumbed, and many biological processes are understood only imperfectly. New tools and computational capabilities are improving biologists' ability to study complex phenomena. Tying together the results of research in the many diverse areas of biology requires a robust theoretical and conceptual framework, upon which a broad and diverse research portfolio of basic biological investigations can be based. The impact of biology on society could be enhanced if discovery and experimentation are complemented by efforts to continuously enrich biology's fundamental theoretical and conceptual basis.

Recommendation 1

Theory, as an important but underappreciated component of biology, should be given a measure of attention commensurate with that given other components of biological research (such as observation and experiment). Theoretical approaches to biological problems should be explicitly recognized as an important and integral component of funding agencies' research portfolios. Increased attention to the theoretical and conceptual components of basic biology research has the potential to leverage the results of basic biology research and should be considered as a balance to programs that focus on mission-oriented research.

Finding 2

Biologists in all subdisciplines use theory but rarely recognize the integral and multifaceted role that theory plays in their research and therefore devote little explicit attention to examining their theoretical and conceptual assumptions. Major advances in biological knowledge come about through the interplay of theoretical insights, observations, and key experimental results and by improvements in technology that make new observations, experiments, and insights possible. The fragmentation of biology into many subdisciplines means both that the mix of these components can differ dramatically from one area to another and that the development of theoretical insights that cut across subdisciplines can be difficult. It is the committee's opinion that all subdisciplines of biology would benefit from an explicit examination of the theoretical and conceptual framework that characterizes their discipline.

Recommendation 2

Biology research funding portfolios should embrace an integrated variety of approaches, including theory along with experiment, observation, and tool development. Biologists in all subdisciplines should be encouraged to examine the theoretical and conceptual framework that underlies their work and identify areas where theoretical advances would most likely lead to breakthroughs in our understanding of life. Workshops sponsored by funding agencies or scientific societies would be one way to facilitate such discussions. The theoretical and conceptual needs identified by such subdisciplinary workshops should then be integrated into the funding programs for those subdisciplines. It would also be worthwhile to sponsor interdisciplinary workshops to identify theoretical and conceptual approaches that would benefit several subdisciplines.

Finding 3

New ways of looking at the natural world often face difficulty in acceptance. Challenges to long-held theories and concepts are likely to be held to a higher standard of evidence than more conventional proposals. Proposals that break new ground can face difficulty in attracting funding, for example those that cross traditional subdisciplinary boundaries, take a purely theoretical approach, or have the potential to destabilize a field by challenging conventional wisdom. Such proposals are likely to be perceived as "high-risk" in that they are likely to fail. However, their potential for high impact warrants special attention. Successfully determining which of them deserve funding will require input from an unusually diverse group of reviewers.

Recommendation 3

Some portion of the basic research budget should be devoted to supporting proposals that are high risk and do not fall obviously into present funding frameworks. One possibility is to initiate a program specifically for such "high-risk/high-impact" proposals—whether they are purely theoretical, cross-disciplinary, or unconventional. Another is to encourage program officers to include some proportion of such proposals in their portfolios. A third is to provide unrestricted support to individuals or teams of scientists who have been identified as particu-

larly innovative. Evaluation of these proposals should be carefully designed to ensure that reviewers with the requisite technical, disciplinary, and theoretical expertise are involved and that they are aware of the goal of supporting potentially consensus-changing research. Proposals that challenge conventional theory require not only that the originality and soundness of the theoretical approach be evaluated but also that the biological data being used are appropriate and the question being asked is significant.

Finding 4

Technological advances in arrays, high-throughput sequencing, remote sensing, miniaturization, wireless communication, high-resolution imaging, and other areas, combined with increasingly powerful computing resources and data analysis techniques, are dramatically expanding biologists' observational, experimental, and quantitative capabilities. Questions can be asked, and answered, that were well beyond our grasp only a few years ago. It is the committee's contention that an increased focus on the theoretical and conceptual basis of biology will lead to the identification of even more complex and interesting questions and will help biologists conceive of crucial experiments that cannot yet be conducted. Biologists' theoretical framework profoundly affects which tools and techniques they use in their work. All too frequently, experimental and observational horizons are unconsciously limited by the technology that is currently available. Advances in technology and computing can provide biologists with many new opportunities for experimentation and observation.

For many of the multiscale questions raised in this report, there is a strong need for teams of biologists, engineers, physicists, statisticians, and others to work together to solve cross-disciplinary problems. The interaction and collaboration of biologists with physicists, engineers, computer scientists, mathematicians, and software designers can lead to a dynamic cycle of developing new tools specifically to answer new questions rather than limiting questions to those that can be addressed with current technology. The growing role and shortening life cycle of technology mean that biologists will have to become ever more adept in the use of new equipment and analysis techniques. Understanding the capabilities, and especially the limitations, of new instruments so that experiments are designed properly and results interpreted appropriately will be important in more and more areas of biology.

Because the potential benefits of more precise and rapid measurements of biological phenomena are so high, it will be important for biologists to be aware of both instrumentation capabilities in the physical and engineering sciences and theoretical advances in physics, chemistry, and mathematics that could be integrated into biological research. Conversely, if researchers outside biology are aware of the kinds of questions biologists are now asking, they can use their techniques, instruments, and approaches to advance biological research. Close collaboration between biologists and researchers in other fields has great promise for leveraging the value of discoveries and theoretical insights arising from basic biological research.

Recommendation 4

In order to gain the greatest possible benefit both from discoveries in the biological sciences and from new technological capabilities, biologists should look for opportunities to work with engineers, physical scientists, and others. Funding agencies should consider sponsoring interdisciplinary workshops focused on major questions or challenges (such as understanding the consequences of climate change, addressing needs for clean water, sustainable agriculture, or pollution remediation) to allow biologists, scientists from other disciplines, and engineers to learn from each other and identify collaborative opportunities. Such workshops should be designed to consider not just what is possible with current technology but also what experiments or observations could be done if technology were not an obstacle. Opportunities for biologists to learn about new instrumentation and to interact with technology developers to create new tools should be strongly supported. One possible approach would be the creation of an integrative institute focused on bioinstrumentation, where biologists could work in interdisciplinary teams to conceive of and develop new instrumentation. The National Center for Ecological Analysis and Synthesis and the National Evolutionary Synthesis Center could serve as models for the development of such an institute.

Finding 5

To get the most out of large and diverse data sets, these will need to be accessible and biologists will have to learn how to use them. While technology is making it increasingly cost-effective to collect huge volumes of data, the process of extracting meaningful conclusions from those data remains difficult, time-consuming, and expensive. Theoretical approaches show great promise for identifying patterns and testing hypotheses in large data sets. It is increasingly likely that data collected for one purpose will have relevance for other researchers. Therefore, the value of the data collected will be multiplied if the data are accessible, organized, and annotated in a standardized way. While it is somewhat new to many areas of biology, other fields—like astronomy and seismology—that create massive data sets rely on theory to guide pattern detection and to direct *in silico* experimentation and modeling. Getting the most out of the extensive biological data that can now be collected will increasingly require that biologists broadly develop those skills and collaborate with mathematicians, computer scientists, statisticians, and others. This process of building community databases is well underway in many areas of biology, genomics being a prominent example, but the specialized databases developed by one research community may be unknown or inaccessible to researchers in other fields. Significant resources are needed to maintain, curate, and interconnect biological databases.

Recommendation 5

Attention should be devoted to ensuring that biological data sets are stored and curated to be accessible to the widest possible population of researchers. In many cases, this will require standardization. Providing opportunities for biologists to learn from other disciplines that routinely carry out theoretical research on diverse data sets should also be explicitly encouraged.

1

Introduction

THE TANGLED WEB OF BIOLOGICAL SCIENCE

The diverse living things of our world are endlessly fascinating. Living organisms have a profound impact on the physical world of ocean, landscape, and climate, and around us is a multitude of diverse ecosystems that provide us with a livable environment and many valuable resources. There is a vast array of interactions among living things, including those that characterize human society and the relationship between humans and the rest of the living world. The practice of biological science takes many forms, with observation, exploration, and experiment combining in many ways to gather information and test hypotheses. The means by which these practices are actually carried out is profoundly affected by the technologies available, with new tools regularly opening up new realms to experimentation, observation, analysis, and novel conceptual insight. Both biologists and nonbiologists occasionally caricature biology in these terms—a science dedicated to endless observation, collection, and testing, leading to a snowballing accumulation of facts. Life is so complex and science has examined such a small fraction of its diversity that it seems reasonable to think that a great deal more data are needed before unifying theories can emerge that explain life in all its diversity. One goal of this report is to illustrate that we need not, and do not, sit and wait for theory to emerge as the end game of biological research. Theory is already an inextricable thread running throughout the practice of biology, as it is in all science. Biologists choose where to observe, what tool to use, which experiment to do, and how to interpret their results on the basis of a rich theoretical and conceptual framework. Biologists strive to discern patterns, processes, and relationships in order to make sense of the seemingly endless diversity of form and function. Explanatory theories are critical to making sense of what is observed—to order biological phenomena, to explain what is seen and to make predictions, and to guide observation and suggest experimental strategies. Because the living world is so complex, biological theory is also exceptionally rich and varied.

> *Science is facts; just as houses are made of stones, so is science made of facts; but a pile of stones is not a house and a collection of facts is not necessarily science*
>
> —Henri Poincare, French mathematician and physicist (1854-1912)
> (Mackay, 1991)

What makes the house of biology from the pile of stone facts is the theoretical component.

THE ORIGIN OF THIS REPORT

In 1989 the National Research Council released a report entitled *Opportunities in Biology*. Over 400 pages long and four years in the making, the report provides a detailed snapshot of the state of biology at that time. Eleven different panels described the opportunities awaiting the rapidly diversifying field of biology. Reading the report today, the excitement of that time is palpable. Section after section describes new technologies and promises new discoveries. The technologies span many levels, from the molecular—DNA sequencing technology had recently progressed from manual to automatic—to the ecological, as robotic arms and free-ranging robots were dramatically expanding the ability of deep-sea submersibles to survey and sample the ocean floor. Nearly 20 years later, it appears that in many respects the authors of that report underestimated the power of the new technologies they described. In 1989 a total of 15 million nucleotides of DNA sequence had been determined. The latest generation of sequencing machines can sequence more than 100 million nucleotides per day. Satellites allow biologists to examine changes in landscapes on an ever finer scale and to track wildlife remotely, while the World Wide Web allows them to retrieve and share their data instantly.

The productivity of biological research since 1989 has been extraordinary. At the same time, the explosion of new biological information has consequences. Individual scientists can now collect data on a scale and at a level of detail that surpass any individual's capacity to sift through, analyze, and interpret all that can be collected. Ever more sophisticated experimental approaches to deciphering how the endless variety of biological systems function opens up a universe of potential experiments so vast that no number of biologists even with unlimited resources could undertake them all. In fact, so much information is accumulating, on so many different biological systems, that it has become impossible for any one biologist to stay abreast of all the advances being made even within one subfield, much less throughout all of biology. There is a growing sense that the ability to collect such a large amount of data, while welcome, also poses new challenges: Are the data already collected adequately organized and accessible, and how can the constant influx of new data be put to best use? How do we decide what experiments to do, which data to collect? There is tremendous potential that new technologies and computational approaches will allow biologists to ask and answer questions that were unmanageable in the past and that chemically and physically reasonable explanations for many complicated biological phenomena will continue to emerge. It is worth considering whether we have the tools and resources necessary to identify potentially unifying themes or organizing principles. A sequel to the 1989 report examining in that same spirit today's "Opportunities in Biology" could easily require 800 pages and 22 subcommittees and would identify hundreds of exciting potential areas for biological discovery. Continuing on the ever-widening research path illuminated in the 1989 report would no doubt lead to great achievements—the record of biological research over the last 20 years has been impressive. At the same time, this is an opportune moment to take stock of the field of biology and examine whether a different perspective is in order, one that might allow biological science to advance faster and contribute even more effectively to addressing the pressing needs of society.

Study Process

This project was initiated at the request of, and with the sponsorship of, the National Science Foundation. The committee first met to discuss its charge and goals in October 2006 and then held a workshop to gather additional input in December 2006. Subsequent meetings in the spring of 2007 were held to work on report writing.

The committee was charged to identify and examine the concepts and theories that form the foundation for scientific advancement in various areas of biology, including (but not limited to) genes, cells, ecology, and evolution. It was asked to assess which areas are "theory-rich" and which areas need stronger conceptual foundations for substantial advancement and to make recommendations as to the best way to encourage creative, dynamic, and innovative research in biology. Building on these results, the study was to identify major questions to be addressed by 21st-century biology. The project was to focus on basic biology, but not on biomedical applications. Questions that could be considered by the committee included:

- What does it mean to think of biology as a theoretical science?
- Is there a basic set of theories and concepts that are understood by biologists in all subdisciplines?
- How do biological theories form the foundation for scientific advancement?
- Which areas of biology are "theory-rich" and which areas need stronger conceptual foundations for substantial advancement?
- What are the best ways to bring about advances in biology?
- What are the grand challenges in 21st-century biology?
- How can educators ensure that students understand the foundations of biology?

At its first committee meeting, in order to identify common theories and concepts in biology, each committee member was asked to present the theories and concepts underlying his or her particular area of research and address how those theories and concepts might apply across biology in general. If the hope was that the talks would unearth a set of theories in each area of biology, sets that could then be compared to find commonalities and show which areas were particularly "theory-rich" and where theory was notably lacking, the result was quite different. The talks demonstrated that biologists from all subdisciplines base their work on rich theoretical foundations, albeit of very diverse types and mixtures. What became evident was the universality of the complex interaction between current theories, new observations and experimental evidence, and evolving technological capabilities. Those areas in which prevailing theory is being challenged through observation, experiment, and analysis are likely to be where the most interesting biology research is being done. This should not have been a surprise for this is a common phenomenon—the recognition that facts are accumulating that contradict the prevailing theoretical framework often characterizes highly active and exciting research and a field in which important changes are imminent. At its second meeting, the committee invited a diverse group of biologists, focusing especially on researchers in subdisciplines that were not represented on the committee, to give talks discussing the theories and concepts underlying their research. Again, the talks did not identify discrete sets of theories that characterized particular areas of research, with some areas having richer theory sets than others. The two sets of talks convinced the committee that identifying a list of concepts and theories that underlie different areas of biology, as requested in the first line of the Statement of Task, would not accurately represent the role of theory in biology. This is not to say that biology has no foundational theories that are accepted by all biologists; evolution, cell theory, and the role of DNA in inheritance certainly serve that unifying purpose. However, the committee did not find that each subdiscipline of biology has its own more or less well-developed set of foundational theories. The committee's assessment of which areas of biology were "theory-rich" and which areas needed stronger conceptual foundations for substantial advancement concluded that all areas of biology rest on a rich theoretical framework but that the range and types of theories in use were exceptionally diverse.

Despite the difficulty that the committee found in responding literally to the Statement of Task, the committee welcomed the opportunity to explore the integral role that theory plays in biology and to point out the ways that theory contributes to creative, dynamic, and innovative research in biology. The committee then decided to use a set of broad questions with relevance across many subdisciplines of biology to illustrate the role that theory now plays and might play even more prominently in the future. The goal was to choose questions that would illustrate the many connections across biological scales and subdisciplines, not to cover the field comprehensively nor to identify which new areas of research are the most important or promising. Inevitably, this approach precluded covering any area in depth and made it impossible to mention all of the many interesting and innovative areas of current biological research.

Where Do Transformative Insights Come From?

In the history of biology, one can identify many moments when our understanding of the living world was transformed. Some of these transformative moments have resulted from a deep insight that led to a major change in our theoretical framework. Other transformative moments were triggered by a key observation or experimental result, or by the invention of a new tool for making observations or doing experiments. None of these moments came about, though, without complex interaction among the many components that make up the practice of biology. Certainly one of the most transformative moments in biology was Darwin's exposition of the theory of evolution by natural selection. His insights have since inspired and elucidated more than a century of rewarding observation and experimentation, richly demonstrating how the process of evolution has resulted in so many diverse life forms, functions, and patterns. But what made possible the transformative moment that was Darwin's theoretical insight? First, an accumulation of facts (in the form of diverse fossil remains) emerged that were difficult to reconcile with the prevailing theory of a fixed and unchanging collection of species. Second, the collection and organization of hundreds of thousands of samples of biological specimens in the museums of Europe during the 19th century (made possible by improvements in navigational tools and motivated by a desire to catalog the diversity of creation), as well as Darwin's own observations and collection during his famous voyage—in other words, the curiosity-driven collection of data about the living world—provided the raw material that enabled Darwin's theoretical insight.

Another profoundly transformative moment, the elucidation of the structure of DNA, could not have happened before the key technological capability of X-ray diffraction was available. Together with the evidence that DNA was the critical substance that passed from generation to generation and that its four simple components were always found in a consistent ratio, Watson and Crick brought to their efforts a theoretical construct. (They had models of the diffraction patterns that helical molecules should produce.) The physical evidence provided by the X-ray scattering patterns that DNA was a molecular double helix was the final link that tied the theory and all the observations together, suggested molecular mechanisms of replication and inheritance, and gave rise to a transforming era in biology.

It is important to note that the tangle of facts, observation, experiment, theory, and technology has no particular beginning and certainly no end. At different times, one of these components may receive more emphasis, but major advances in modern biology have never been completed without all of them.

Despite the integral role that theory plays throughout the practice of biology, biologists rarely think of themselves as theoretical scientists. Part of the reason is that the word "theory" can be used to mean many different things, ranging from a mere hunch to a set of mathematical equations codifying a "law of nature." Although the word is generally used by scientists more rigorously than the general public to mean an explanatory framework supported by a large body of observational and experimental evidence, even scientists tend to confine the idea of "theoretical science" to the practice of developing mathematical equations to represent a large body of phenomena. While mathematical, computational, and quantitative approaches have important roles in biology, confining the definition of theory to these efforts fails to capture the texture of theory in biology. In Chapter 2 this report adopts a more flexible description of theory as a "family of models" that can be, among other things, physical, visual, verbal, mathematical, statistical, descriptive, or comparative. The models need not even all be entirely consistent with each other (just as it is sometimes useful to model light as a wave, sometimes as a particle), the important characteristic being that the model is a representation of some aspect of nature for the purpose of study. Using this view of biological theory makes it clear how ubiquitous it is in scientific practice. For example, if one's model of the genome suggests that only protein-coding regions are important for development, one may adopt an RNA extraction technique that selects only transcripts with poly-A tails. An alternative model that includes a functional role for noncoding sequences in development would require a different extraction technique. A scientist whose model of cellular robustness rules out the possibility of life below pH 3 or above 90°C will not look for bacteria in the human stomach or in the hot springs of Yellowstone. Explicit recognition that one's theoretical and conceptual framework is affecting choices throughout one's research—from the tools used, to the experiments done, to the interpretation of the results and more—may help stimulate truly innovative and transformative research.

Because theory in biology sometimes corresponds poorly with common stereotypes of theoretical science, biologists and others often fail to recognize its importance. Yet theory is clearly an integral part of the process of biological research and is vital to its success. It is time for biology to take a step back and think carefully about balancing the attention being paid to theory in relation to observation, experiment, and technology development. Would an explicit emphasis on the theoretical and conceptual component of biological research be fruitful, and if so, how would that best be done?

Facilitating Future Advances in Biology: Achieving a Balance

The emergence of a new insight is, by its very nature, unpredictable. In retrospect, however, it is possible in many cases to dissect the relative contribution of theory to the great discoveries of the past. But is it also possible to look at biological research today and determine whether emphasis on one area or another would be most likely to drive innovation and transformation of the field? The topic of transformative research was recently the focus of a National Science Board report, *Enhancing Support for Transformative Research at the National Science Foundation* (May 7, 2007). That report states that "[t]ransformative research is defined as research driven by ideas that have the potential to radically change our understanding of an important existing scientific or engineering concept or leading to the creation of a new paradigm or field of science or engineering. Such research is also characterized by its challenge to current understanding or its pathway to new frontiers." This study's Statement of Task asked the committee to consider whether biology might benefit from an intensive focus on developing theoretical or conceptual foundations: in other words, to consider whether transformative moments would be likely to be driven by a focus on theory.

The increasing fragmentation of the practice of biological research into subdisciplines makes it challenging for biologists to recognize theories that cut across biological scales. The body of knowledge about biological systems has grown so vast so rapidly, and the variety of approaches is now so numerous, that it has become impossible for any one scientist to stay fully abreast of the cutting edge of research—where experiment and observation are actively generating new theories and models (and vice versa)—throughout the full range of biological research. Perhaps even more challenging is the effort to understand enough about other scientific disciplines to know whether the research being done on a specific biological question could inform, advance, or build on research being done outside biology.

Key Questions

The committee chose to illustrate the role theory can play in answering broad questions in the field of biology and addressing grand challenges for society by developing a set of questions that have relevance across many subdisciplines of biology. These questions consider those characteristics that are unique to living systems and are questions that perhaps only the study of living things can answer.

The questions vary. Some focus on characteristics that are similar across many biological scales, while others focus on the incredible diversity of life. Still others take an explicitly theoretical point of view. The committee makes no claim that this set of questions is comprehensive, but simply aims to give a set of important examples of how explicit attention to theory might contribute to answering these kinds of questions—questions that would be difficult to address through a traditional approach. The goal was to choose questions that would illustrate the many connections across biological scales and subdisciplines, not to cover the field comprehensively, nor to identify which new areas of research are the most important or promising. Inevitably, this approach precluded covering any area in depth and made it impossible to include all of the many interesting and innovative areas of current biological research.

The questions are listed below. The summary at the beginning of the report gives a brief overview of each question, and within the body of the report a separate chapter addresses each one.

1. Are there still new life forms to be discovered?
2. What role does life play in the metabolism of planet Earth?
3. How do cells really work?
4. What are the engineering principles of life?
5. What is the information that defines and sustains life?
6. What determines how organisms behave in their worlds?
7. How much can we tell about the past—and predict about the future—by studying life on earth today?