The end of “naïve reductionism”: rise of systems biology or renaissance of physiology?

Kevin Strange

Departments of Anesthesiology, Molecular Physiology and Biophysics, and Pharmacology, Vanderbilt University Medical Center, Nashville, Tennessee

Submitted 7 December 2004; accepted in final form 13 December 2004

Strange, Kevin. The end of “naïve reductionism”: rise of systems biology or renaissance of physiology? Am J Physiol Cell Physiol 288:C968–C974, 2005; doi:10.1152/ajpcell.00598.2004.—Systems biology is an emerging discipline focused on tackling the enormous intellectual and technical challenges associated with translating genome sequence into a comprehensive understanding of how organisms are built and run. Physiology and systems biology share the goal of understanding the integrated function of complex, multicomponent biological systems ranging from interacting proteins that carry out specific tasks to whole organisms. Despite this common ground, physiology as an academic discipline runs the real risk of fading into the background and being superseded organizationally and administratively by systems biology. My goal in this article is to discuss briefly the cornerstones of modern systems biology, specifically functional genomics, nonmammalian model organisms and computational biology, and to emphasize the need to embrace them as essential components of 21st-century physiology departments and research and teaching programs.

functional genomics; model organisms; genetics

MOLECULAR BIOLOGY drove a powerful reductionist or “molecule-centric” approach to physiology research in the last half of the 20th century. Reductionism is the attempt to explain complex phenomena by defining the functional properties of the individual components that compose multicomponent systems. Bloom (8) has referred to the post-genome sequencing era as the end of “naïve reductionism.” Reductionist methods will continue to be an essential element of all biological research efforts, but “naïve reductionism,” the belief that reductionism alone can lead to a complete understanding of living organisms, is not tenable. Organisms are clearly much more than the sum of their parts, and the behavior of complex physiological processes cannot be understood simply by knowing how the parts work in isolation.

Systems biology has emerged in the wake of genome sequencing as the successor to reductionism (42, 49, 50, 72). The “systems” of systems biology are defined over a wide span of complexity ranging from two macromolecules that interact to carry out a specific task to whole organisms. Systems biology is integrative and seeks to understand and predict the behavior or “emergent” properties of complex, multicomponent biological processes. A systems level characterization of a biological process addresses three main questions. First, what are the parts of the system (i.e., the genes and the proteins they encode)? Second, how do the parts work? Third, how do the parts work together to accomplish a task?

Physiology is defined as “the study of living organisms and their parts.” The discipline of physiology is inherently integrative. This integrative nature is reflected in the enormous breadth of physiological research, which ranges from the study of single molecules to molecular pathways, cells, tissues, organs, and whole organisms. Physiology and systems biology thus share the goal of understanding the integrated function of complex, multicomponent biological systems. In my opinion, physiology and systems biology are synonymous.

If physiology and systems biology share the same overall scientific goal, then why is systems biology on a rapid rise, whereas physiology is, as noted recently by Caplan (14), often viewed as a “quaint and vaguely anachronistic” discipline? Systems biology arose to tackle the enormous intellectual and technical challenges associated with translating genome sequence into a comprehensive understanding of how organisms are built and run. Systems biologists characterize complex physiological processes by utilizing a variety of tools including large-scale functional genomics methods, genetic analysis of nonmammalian model organisms, bioinformatics, and computer modeling. These same tools and experimental approaches have not been widely exploited or promoted by physiologists. Systems biologists focus on identifying and understanding in an integrated fashion the network of functionally interacting proteins that give rise to biological processes. Physiology research over the past 30 years has tended to move away from integration and synthesis toward ever-increasingly detailed biophysical analysis of proteins or whole animal physiological systems. My goal in this article is to discuss briefly the cornerstones of modern systems biology, specifically functional genomics, nonmammalian model organisms and computational biology, and to emphasize the need to embrace them as essential components of 21st century physiology departments and research and teaching programs.

FUNCTIONAL GENOMICS: FISHING ON A DIFFERENT SCALE

Developing a systems level understanding of a physiological process requires identification of the genes, and the proteins they encode (i.e., the “parts”), that work together to give rise to that process. Functional genomics is therefore an essential foundation of systems biology research. As a discipline or field, functional genomics is one that develops and utilizes large-scale and high-throughput methodologies to define and analyze gene function at a global level (35, 80, 106). Functional genomics is also inherently integrative and attempts to generate insights into gene function by integrating data ob-
tained from multiple large-scale datasets (e.g., Refs. 4, 29, 33, 80, 101, 106).

The tools of functional genomics have not been widely accepted in the physiology community and it is not uncommon to hear criticisms about their development and use. Large-scale data collection methods are often viewed as “fishing expeditions” that are not hypothesis driven and that fail to provide definitive mechanistic insights into biological processes. Obviously, hypotheses cannot be formulated and hypothesis-driven experiments cannot be carried out without first making observations or “fishing.” Physiologists employ a vast array of approaches for making observations, ranging from X-ray crystallography, quantitative microscopy, and patch-clamp electrophysiology to measurements of whole animal behavior. These experimental strategies differ from global analyses of gene function only in scale. All of these approaches are equally capable of providing mechanistic insights and the foundations for hypothesis-driven experimental studies. The following section briefly describes the power of large-scale data collection methods for identifying networks of genes that underlie complex physiological processes.

DNA microarrays and related technology were one of the earliest functional genomics tools developed for large-scale analysis of gene function. To the casual observer, microarrays are often viewed as little more than multigene Northern blots. In reality, microarray experiments can provide unique insights into the function of uncharacterized genes and can identify groups of interacting genes that give rise to a physiological process of interest.

The challenge in a microarray experiment is how to detect meaningful patterns in the massive amounts of data generated and how to deduce experimentally testable functions from those patterns (55, 76, 84). Several approaches have been developed to address this problem. For example, comparison of reference databases or “compendia” of gene expression profiles generated under different physiological and developmental conditions, in different genetic backgrounds, and in different cell types and tissues of an organism as well as across multiple, evolutionarily diverse species allows identification of networks of coregulated genes that carry out specific tasks. Microarray analyses can also provide insights into mechanisms of gene regulation, evolution and the etiology of disease (e.g., 6, 39, 40, 48, 59, 87, 88).

Reverse genetics attempts to identify the function of a gene by disrupting its coding region or by decreasing the levels of mRNA and protein and analyzing the resultant phenotype. Physiologists utilize knockout and knockdown strategies to study the function of single genes. However, it is now possible using double stranded RNA (dsRNA)-mediated gene interference (RNAi) to carry out large-scale and whole genome reverse genetic analysis (see Refs. 21, 34, 74, and 93 for reviews on the mechanism of RNAi). The nematode Caenorhabditis elegans has proven to be particularly amenable to whole genome RNAi screening. RNAi can be induced in worms by injecting them with dsRNA (24), by soaking them in dsRNA solutions (92), or by feeding them bacteria producing dsRNA (46, 94, 95). When worms are fed dsRNA-producing bacteria or soaked in dsRNA solutions, the dsRNA is absorbed across the intestinal epithelium and then spreads systemically to the animal’s somatic cells and germline. In cultured C. elegans cells, RNAi is triggered simply by the addition of dsRNA to the culture medium (15).

Recently, Kamath et al. (45) generated a reusable RNAi library (now available from MRC geneservice) consisting of 16,757 bacterial strains, each of which expresses a unique dsRNA. These dsRNAs correspond to ~86% of the predicted genes in the worm genome. To date, large-scale RNAi feeding has been used to identify genes involved in fat metabolism (3), aging (54, 65), early embryonic development (111), transposon silencing (99), protection of the genome against mutations (75), osmotic stress resistance (53), and regulators of polyglutamine aggregation (67).

Clemens et al. (17; see also Ref. 105) demonstrated that dsRNA added to the culture medium induces RNAi in S2 and other Drosophila cell lines. This seminal observation laid the foundation for large-scale RNAi screening in cultured Drosophila cells to identify genes important for specific cellular phenotypes and physiological processes including phagocytosis (78), cell shape regulation (47), cell growth and viability (10), innate immunity (25), Hedgehog signaling (57), and regulation of alternative splicing (71). Recently, RNAi living-cell microarrays were developed and have been used to identify genes that regulate cell growth and Akt phosphorylation (103). A Drosophila genome cDNA library suitable for automated dsRNA synthesis is available from MRC geneservice.

Gene silencing by RNAi in mammalian cells is considerably more technically complicated than that in invertebrates (see Refs. 32 and 60 for a review on the use of small interfering RNAs, or siRNAs, in mammalian cells). However, despite the inherent experimental difficulties, large-scale RNAi screening in mammalian cells has been undertaken and will likely become a routine experimental approach in the near future as whole genome mammalian siRNA libraries and high-throughput methodologies are developed and refined (7, 64, 70, 81, 83, 109).

Protein-protein interactions play critical roles in most biological processes. With the availability of complete genome sequences, numerous efforts are being undertaken to map protein-protein interactions on a genome-wide scale (26, 100). Large-scale protein interaction screening has been carried out in yeast (43, 97), C. elegans (9, 19, 19, 56, 102), and Drosophila (31) using high-throughput yeast two-hybrid analysis. Interaction screens have also been undertaken in yeast using affinity purification and mass spectrometry to identify components of isolated protein complexes (28, 36). This latter approach has the advantage of being able to identify complexes containing three or more proteins.

Large-scale and genome-wide protein interaction maps can lead to new understanding of gene function in several ways. First, biological insights can be gained into proteins with unknown functions by linking them to proteins that operate in characterized biological pathways. In addition, protein-protein interaction mapping may identify new interactions and hence functions for previously characterized proteins and they can identify networks of interactions that are required for complex biological processes, such as the DNA damage response (5, 9); kinetochore function and regulation (82); DNA helicase (66), spliceosome (26), and proteosome function (13, 19); stress response signaling pathways (12); and tissue development (102).
In addition to transcriptional profiling, reverse genetic screening and protein-protein interaction mapping, many other genome-wide and large-scale functional genomics projects are underway. For example, efforts are ongoing in yeast to quantify the levels and identify the cellular location of every expressed protein (20, 30, 41, 79). Previous studies (63) have carried out large-scale localization of proteins with green fluorescent protein tagging in *Drosophila* (63). Numerous efforts are underway to generate mutants and knockout gene expression on a genome-wide scale in various organisms, including yeast, plants, *C. elegans*, *Drosophila*, and mouse (reviewed in 11, 18, 77, 85, 98). Large-scale analysis of protein function has been initiated using a variety of approaches, including peptide and protein arrays and glutathione-S-transferase fusion methods (reviewed in Ref. 110). Mass spectrometry and NMR are being used to carry out large-scale analysis and quantification of metabolite levels in organisms and cells (reviewed in Refs. 69, 73, 96). All of these large-scale projects, particularly when their data sets are integrated (e.g., 29, 33, 44), provide foundations for identifying interacting gene networks and hypothesis-driven investigations.

**DECIPHERING GENETIC CODE: THE POWER OF NONMAMMALIAN MODEL ORGANISMS**

The physiologist and Nobel laureate August Krogh believed that there is an ideal organism in which almost every physiological problem could be studied most readily, a belief often referred to as the “Krogh Principle” (51, 52). If an investigator’s goal were a systems-level characterization of a physiological process, the optimal model organism for him/her to utilize would be one that is genetically and molecularly tractable; in other words, an organism in which forward and reverse genetic analyses could be carried out readily, rapidly, and economically.

Model organisms such as *Escherichia coli*, *Saccharomyces*, *C. elegans*, *Drosophila*, zebrafish, and the plant *Arabidopsis* have become cornerstones of systems biology research. They have been likened to the Rosetta stone (42), which provided modern scholars the tools needed to decipher Egyptian hieroglyphics. Similarly, nonmammalian model organisms provide physiologists the experimental tools necessary to decipher the genetic code that underlies complex physiological processes common to all life.

*C. elegans* provides a particularly striking example of the experimental utility of nonmammalian model organisms (see Ref. 86 for a review). Worms have a short life cycle (2 to 3 days at 25°C), produce large numbers of offspring by sexual reproduction and can be cultured easily and inexpensively in the laboratory. Sexual reproduction occurs by self-fertilization in hermaphrodites or by mating with males. Self-fertilization allows homozygous animals to breed true and greatly facilitates the isolation and maintenance of mutant strains. It is also a useful feature if mutant worms are paralyzed or uncoordinated because reproduction does not require movement to find and mate with a male. Mating with males, however, is essential for moving mutations between strains. The reproductive and laboratory culture characteristics of *C. elegans* make it an exceptionally powerful model system for forward genetic analysis. Mutagenesis and genetic screening allows identification of the network of genes underlying a physiological process of interest and can also provide important and novel mechanistic insights into the molecular structure and function of proteins.

In addition to forward genetic tractability, *C. elegans* also has a fully sequenced and well-annotated genome. Genomic sequence and virtually all other biological data on this organism are assembled in readily accessible public databases (e.g., WormBase; http://www.wormbase.org). Numerous reagents including mutant worm strains and cosmids and YAC clones spanning the genome are freely available through public resources. Creation of transgenic worms is relatively easy, inexpensive, and rapid, requiring little more than an injection of transgenes into the animal’s gonad or bombardment with DNA-coated microparticles. As noted earlier, *C. elegans* gene expression can be specifically and potently targeted for knockdown using RNAi, either at the single worm level by injection of double-stranded RNA (dsRNA), or at the population level by feeding worms dsRNA-producing bacteria. Finally, *C. elegans* is a highly differentiated animal but is composed of <1,000 somatic cells. This relatively simple anatomy greatly facilitates the study of physiological processes and has made it possible to trace the lineage of every adult cell beginning with the first cell division (89, 90), and to generate a complete wiring diagram of the 302 neuron adult hermaphrodite nervous system (104).

Despite their obvious impact on biomedical research, physiologists often criticize nonmammalian model organisms such as *C. elegans* as being too “simple” or too evolutionarily ancient to provide significant insight into mammalian physiology. However, biology is rarely reinvented by natural selection. Instead, natural selection takes a working plan and modifies it. Addressing complex physiological problems in a less complex model organism that is genomically defined and genetically tractable and where it is more straightforward and economical to manipulate gene function is a rational and powerful experimental strategy.

A complete discussion of the unique insights and understanding of basic biological processes provided by nonmammalian model organisms is well outside the scope of this article. However, a brief mention of a few examples relevant to common problems under investigation in many physiology departments is warranted. For example, genetic screens of phototransduction in *Drosophila* led to the discovery of the first “transient receptor potential” or TRP channel (61). *Drosophila* TRPs in turn led to the discovery of TRP channels in mammals and an explosion of research that has shed new light on sensory physiology and cellular signaling (16, 62). The ability to detect and respond to mechanical force (i.e., mechanotransduction) underlies physiological processes such as hearing, proprioception, and blood pressure regulation. *C. elegans* has provided the most comprehensive understanding of how cells and whole organisms detect mechanical force (91). The demonstration that degenerin/epithelial Na⁺ channels (ENaCs) play critical roles in worm mechanotransduction led to extensive research that subsequently linked this channel family to mechanotransduction in mammals (91). Cystic fibrosis is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR). The most common disease-causing allele is ΔF508, which results in protein misfolding and degradation (2). Protein trafficking is highly conserved in yeast and mammals (22). Several investigators have recently begun exploiting the genetic tools of yeast to define the mechanisms...
and network of proteins involved in CFTR biogenesis and degradation (27, 107). Lipid metabolism and digestive physiology are being genetically characterized in zebrafish using fluorescent phospholipid reporters (23). In addition to defining the genetic bases of basic physiological processes, nonmammalian model organisms from yeast to zebrafish are providing invaluable understanding of the pathophysiology associated with diabetes; obesity; cancer; heart, muscle, blood, developmental, and neurological disorders; neurodegeneration; aging; infection; kidney disease; addiction; and immune system dysfunction (e.g., 1, 58, 68).

**BIOINFORMATICS AND COMPUTATIONAL BIOLOGY: SEEING THE FOREST THROUGH THE TREES**

The discipline of bioinformatics was spawned by the massive amounts of data generated by genome sequencing and functional genomics. Bioinformatics merges biology with computer science, statistics and information technology to develop algorithms and statistical methods necessary to manage, access, and analyze large data sets and assess their interrelationships. These tools reveal functional information in genome scale data and thus provide the foundation for formulating experimentally testable hypotheses.

Computational biology, by modeling system architecture, information flow, and information processing, provides a predictive understanding of how a system responds to perturbations. Systems modeling of a physiological process beginning at the level of genes and gene networks is a highly iterative process involving cycles of data collection, quantitative modeling, hypothesis formulation and testing, and model refinement. Modeling efforts can provide unique biological insights that could not have been obtained by intuition and experiment alone. The development of an integrated system level model that fully predicts the behavior of a system under all biologically relevant conditions implies that an in-depth understanding of how the system works has been achieved.

In addition to biological insight and understanding, system models offer enormous promise for treating disease. Imagine, for example, that one has identified the network of genes underlying a physiological process such as a developmental circuit, the electrical activity of an excitable cell, a signaling pathway, or solute and fluid transport in a lung epithelial cell, that the interactions and functions of those genes have been defined, and that system architecture and behavior have been modeled. A model that is a realistic representation of the physiological process allows prediction of how pathophysiological perturbations such as a gene mutation or a pathogen will affect system behavior. If pathophysiological behavior can be predicted, then it should be possible to predict how the system can be counterperturbed to correct abnormal function. Computational biology is thus essential for rational identification of drug and therapeutic targets and can aid in drug design by predicting side effects (37, 38).

**THE POSTGENOME PHYSIOLOGY DEPARTMENT**

I believe that physiology should be viewed as a leader in systems biology research and teaching, not as a “quaint and vaguely anachronistic” discipline. However, what needs to be done to move into a leadership position? First, we must reinvigorate and fully promote integrative research and teaching, which are foundations and unique strengths of physiology. Unfortunately, integrative thinking and synthesis often get lost in the drive to understand proteins at deeper and deeper levels. We must emphasize to our graduate and postdoctoral students the importance of seeing and understanding the “big picture” through the parts of the puzzle that their research focuses on. We must teach them to synthesize instead of compartmentalize, to transform an understanding of molecular function into an understanding of molecular, cell, tissue, organ, and whole animal integrative physiology. The courses we teach graduate students, particularly students in interdisciplinary Ph.D. programs, should emphasize synthesis and integration, and stress that integrative research, physiology, and systems biology are foundations of 21st century biomedical research. Along with our students, we must also educate our university administrators on the critical importance of physiology and integrative research, and we must make them fully aware that they are key components of the NIH Roadmap (108).

Physiologists utilize an extraordinary array of powerful research tools. However, genetics and functional genomics are poorly represented, if at all, in most physiology departments. The faculty of a 21st century physiology department should be composed of a “critical mass” of investigators who are defining the genetic bases of physiological processes using state-of-the-art genome scale experimental tools and genetic analysis of genetically tractable model organisms such as yeast, *C. elegans*, *Drosophila*, zebrafish, and even bacteria and plants. These investigators should be working side-by-side with physiologists using classical vertebrate and mammalian models. Close interactions will facilitate the use of functional genomics tools by mammalian physiologists and will enhance the transfer of genetic understanding of nonmammalian model organism biology into a deeper molecular understanding of mammalian physiology. It will also facilitate the development of new genetic models for investigating physiological and pathophysiological problems studied previously only in mammalian experimental systems.

Mathematical modeling has a long and rich history in physiology. However, despite its power, modeling is not the norm in physiological research. However, as we move inexorably away from pure reductionism and toward prediction of complex systems behavior, mathematical modeling becomes essential. The information flow in even the simplest systems is likely too great for understanding through intuition and experiment alone. Modeling will thus be an integral component of post-genome physiological research programs.

The vast majority of biomedical scientists will not be able to develop systems models on their own. This means that physiologists will more than ever need to engage in interdisciplinary research collaborations and partner with mathematicians, engineers, computer scientists, and physicists. Enormous obstacles need to be overcome for such collaborations to bear fruit. At a minimum, physical scientists will need to learn the language of biology and biologists will need to learn the language of mathematics and computer modeling. The territorial walls that surround many academic departments will have to be replaced with interdisciplinary research programs, such as centers and institutes. Physiology departments should become leaders in developing interdisciplinary research and training programs that are focused on systems-level analysis of physiological processes.

---

*AJP-Cell Physiol.* • VOL 288 • MAY 2005 • www.ajpcell.org
In conclusion, with the availability of complete genome sequences, an integrative understanding of whole organisms from the level of single genes and gene networks is now at hand. However, given our current level of understanding, a sequenced genome represents little more than lines of code (i.e., genes) that specify how to synthesize proteins. Genome sequence must be deciphered into a set of instructions that allow us to understand when and where proteins are synthesized, when and how groups of proteins are assembled into functional complexes and pathways that accomplish specific tasks, and when and how tasks are assembled to build and run organelles, cells, tissues and organs, and the whole organism. Enormous intellectual and technical challenges must be overcome to achieve this goal. The discipline of systems biology has arisen in the post-genome era to tackle these challenges. Physiology and systems biology share the goal of understanding the integrated function of complex systems from the level of genes to the whole organism. Yet physiology as an academic discipline runs the real risk of fading into the background and being superseded organizationally and administratively by systems biology. To avoid this fate, physiology departments and physiologists must fully embrace the cornerstones of systems biology. To prevent this, physiology departments and physiologists must focus on the cornerstones of systems biology research including functional genomics, genomics, non-mammalian model organisms, computational biology, and interdisciplinary research efforts. We must also move away from naïve reductionism and reemphasize in our research and teaching the central importance of integration and synthesis. If we successfully accomplish these goals, the post-genome sequencing era can become the era of a physiological renaissance.

ACKNOWLEDGMENTS

The author thanks Villu Mariq, Bob Putnam, Jerod Denton, Mike Caplan, and Mark Zeidel for critically reading the manuscript and for helpful comments and insight.

GRANTS

This work was supported by National Institutes of Health Grants R01 DK-51610, R01 DK-61168, P01 DK-58212, and R21 DK-64743.

REFERENCES


