BIOINFORMATICS

What lies beyond bioinformatics?

Bernhard O. Palsson

Vast amounts of basic genetic and biochemical information are rapidly becoming available. Sequencing technology is providing us with complete information about the genetic makeup of simple cells, and more complex organisms will soon follow. Commensurately, there are now well over 100 biological databases available on the World Wide Web that contain information about the genetic makeup and biochemical characteristics of a variety of cells and cellular processes[1,2]. Links between individual databases[3,4] are resulting in essentially complete genetic and metabolic information about such specific bacterial cells as *Escherichia coli* and *Haemophilus influenzae*. Simultaneously, genetic and regulatory similarities between such popular model organisms in developmental biology as yeast, worms, flies, and mice are being unraveled[5]. And thus enters the field of bioinformatics.

With all this information about the underlying molecular determinants of living systems becoming available at a truly astounding rate, we are faced with the question of where it will all lead. Astronomy, physics, chemistry, and engineering have each undergone similar phases in their histories. When large amounts of data become available, they are organized, analyzed for regularities (which eventually become "laws" and "principles"), and these rules in turn are used to build scientific theories and to implement practical applications. For example, detailed knowledge of the laws of aerodynamics allows the training of pilots in simulators, and the design of modern aircraft now proceeds directly from a computer workstation, with minimal, if any, prototyping. We should, therefore, expect that we are on the threshold of a new era in the biological sciences, an era in which a new molecularly based theoretical

biology will emerge. As with other applied fields of science, the dictum "nothing is more practical than a good theory" will eventually apply to cell and molecular biology. Making a statement such as this seriously was unthinkable only a few years ago.

How is this process likely to unfold? Initially, one would expect that analysis of the
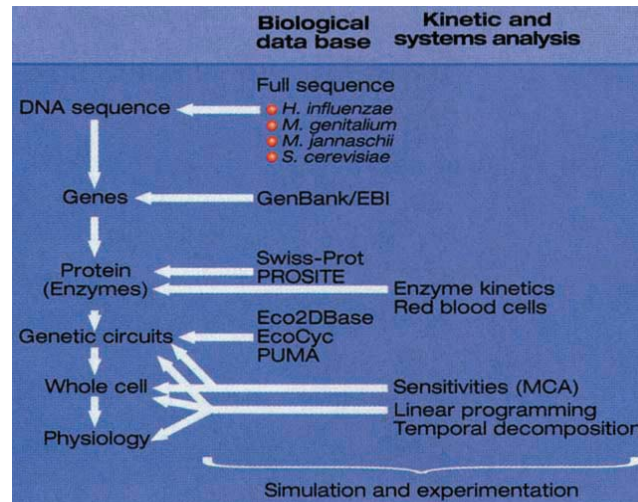

Figure 1. A schematic illustration of the multistep relationship between genetics and physiology. The first few steps are being extensively characterized in genetic and biochemical terms, and some of the databases available are indicated. However, the system behavior of multiprotein systems, i.e. genetic circuits, will require kinetic and systems analysis. Some of the available methods are indicated and discussed in the text. The development of these last steps is now becoming critical to the attainment of the overall genetics–physiology relationship.

Table 1. Some characteristics of genetic circuits and the analysis methods required to understand them.

Characteristic	Analysis method
They are complex	Bioinformatics
They are autonomous	Control theory
They are robust	System science
They function to execute	Transport and kinetic theory, a physicochemical process
They have "creative functions"	Bifurcation analysis
They are conserved, but can adjust	Evolutionary dynamics

complete genome sequences grows, many more such studies will appear, deepening our understanding of basic biological processes. The base pair sequence in the human genome amounts to about one gigabyte of information, equivalent to the amount of information that many of the readers of this article store on their personal computers. We can anticipate that the use of model organisms will lead to functional assignment of most of the 70,000–100,000 human genes, and reduce their roles into much fewer multicomponent cellular functions.

How will the reduction of gene number to much fewer cellular and physiological functions take place? The relationship between genetics and physiology has many layers, as illustrated in Figure 1. Gene sequences allow the identification of open reading frames (ORFs). The base pair sequence of the ORFs in turn allows for the functional assignment of the defined gene. Although not always unambiguous, such assignments are being carried out with increasing accuracy. Sequence is important, and so is the functional assignment. However, the interrelatedness of genes may prove to be even more important. Establishing these relations and studying their systemic characteristics now become urgent. The vast majority of cellular functions rely on the coordinated action of the products from multiple genes. Such coordinated function can be viewed as a genetic circuit, representing cellular "wiring diagrams"—the collection of different gene products that together are required to execute a particular function. The functions of such genetic circuits are diverse, including such functions as DNA replication, translation, the conversion of glucose to pyruvate, the laying down of the basic body plan of multicellular organisms, and cell motion. It is likely that we will come to view cellular functions within this framework, and the physiological function of cells and organisms as the coordinated functions of multiple genetic circuits. Consequently, we will need to develop a conceptual framework within which to describe and analyze these circuits.

Not all the properties of genetic circuits are clear at present, but some are summarized

Bernhard O. Palsson is a professor at the department of bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412 (palsson@ucsd.edu).

genetic and biochemical information will result in new knowledge and insight into cell and tissue function. Indeed, just a year after the publication of the first two complete bacterial genomes, comparative analysis has led to the suggestion that there are 256 genes, which together perform about a dozen cellular functions and which constitute a minimal gene set for a modern cell[6]. As the number of

COMMENTARY

in Table 1. For many of the characteristics shown, it is clear what methodologies are needed to describe and analyze them. Genetic circuits tend to have many components; they are complex. From the standpoint of system science, they are "robust," i.e., one can often remove their components without compromising their overall function. For instance, many knockout mice have normal phenotypes, even if the genes removed were thought to have critical roles. Furthermore, multiple failures in cell cycle regulation are needed for a transformation to a malignant phenotype.

Genetic circuits have built-in controls. That is, once expressed, their function is autonomous. They are capable of functioning by themselves. Studies of the characteristics of systemic regulation of metabolic pathways are perhaps the furthest advanced. For instance, time-scale hierarchy plays a role in the stabilization of metabolic function[7], and temporal decomposition of metabolic dynamics has helped elucidate the metabolic blueprint that underlies metabolic function in the human red blood cell[8]. Embedded in these control structures are the capabilities to perform what has been called "creative functions"[9]. Such functions include sustained oscillatory behavior, and multiple steady states[7] leading to built-in "decision making" mechanisms. Some of these have been experimentally demonstrated[10]. Such kinetic characteristics are studied by bifurcation theory[7].

Many genetic circuits carry out a physicochemical process. For instance, some cell-to-cell communications involve extracellular events (diffusion of signaling molecule) and subsequent intracellular events (kinetic behavior of signal transduction cascades). Chemical kinetic theory and transport phenomena will be used to describe these characteristics[11]. Finally, it appears that once a genetic circuit has been established, it is evolutionarily preserved. This preservation leads to unity in biology, such as the universal glycolytic and Ras-signaling pathways. However, such genetic circuits change over the course of evolution, and, with increasing organism complexity, they must coordinate their activities with new circuits.

The systemic description of several complex genetic circuits has already appeared, including the λ bacteriophage[12], the eukaryotic cell cycle[13], and red cell metabolism[14]. Methods of systems science and kinetic theory will inevitably play a role, as biological function is dynamic and systemic. Analysis methods, such as metabolic control analysis (MCA)[15], flux-balance analysis[16], and modal analysis[8] have proven useful for metabolic studies.

Some of these existing methods will be important in analyzing proliferating genomic databases. For instance, flux-balance analysis is based only on metabolic stoichiometry and

biosynthetic demands—its success is dependent on detailed knowledge of the biochemical function of metabolic enzymes. The assignment of ORFs will immediately give the "metabolic genotype"[16] of a freshly sequenced genome. The metabolic capabilities of the metabolic genotype can then be assessed and
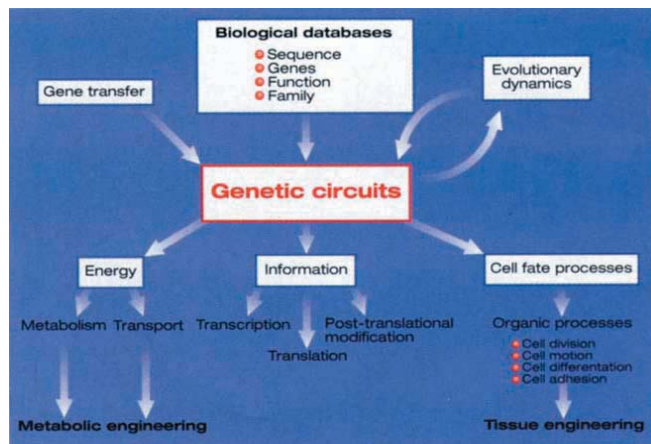

Figure 2. Genetic circuits. The definition of genetic circuits will most likely arise from bioinformatics. Genetic circuits will need a classification system: Major categories are indicated, in particular to show that some underlie important metabolic and tissue engineering applications of cell and molecular biology.

compared with other metabolic genotypes by using this analysis method. Such studies will help us understand relative metabolic capabilities of different cells and guide the development of industrial strains. Although at present enzyme kinetics properties are not as well catalogued and "assignable" as ORFs, it is likely that bioinformatics will eventually allow similar analysis of metabolic dynamics. We can anticipate the use of existing methods to analyze the systemic kinetic behavior of newly sequenced genomes and their metabolic genotypes.

Perhaps the best available methods for systems analysis in biology are those for metabolism. The details of metabolism were mostly worked out by the 1950s, beginning with glycolysis in the 1930s. Thus, when computer capabilities developed, the systemic properties of metabolism could be studied. One can make the argument that the formulation of the signal transduction pathways that govern mammalian cell and tissue behavior is in a state similar to that of metabolic research in the early 1940s. The genes and their individual functions are becoming known, but we are in the midst of determining how they relate to one another. However, we can expect that the elucidation of these pathways and their classification into genetic circuits will proceed at a much faster pace. The systems analysis methods developed for metabolism may prove applicable to these circuits, but undoubtedly new methods will need to be developed.

Accepting the concept of a genetic circuit seems straightforward. However, the implica-

tions of this acceptance are quite profound. We will view bioinformatics as a way to establish, classify, and cross-species correlate genetic circuits. The beginning of such classification is illustrated in Figure 2. Metabolism, information processing, and cellular fate processes represent some of the major categories of genetic circuits.

Gene therapy may no longer be viewed as replacing a "bad" gene, but fixing a "malfunctioning" genetic circuit. Evolution may be viewed as the "tuning" or "honing" of circuits to improve performance and chances of survival. Classifying organisms based on the types of genetic circuits they possess may lead to "genomic taxonomy." Ex vivo evolutionary procedures for adjusting circuit performance for human-specified functionalities will emerge. Understanding the function of genetic circuits is fundamental to applied biology, in fields as diverse as metabolic engineering and tissue engineering.

The dynamics of paradigm shifts in the history of science and scientific revolutions have been studied, most notably by Thomas Kuhn. It seems clear that as a result of bioinformatics, biology is currently undergoing multiple changes of its conceptual framework. I have suggested here that the genetic circuit concept will become an important new biological paradigm, and will become fundamental to our treatment of the relationship between genetics and physiology.

1. Robinson, C. (ed.). 1996. *Trends Biotechnol.* **14**:261–321.
2. Walker, R.T. et al. (eds.). 1996. *Nucleic Acids Res.* **24**:1–252.
3. Karp, P.D. 1996 *Trends Biotechnol.* **14**:273–279.
4. Karp, P.D. 1996. *Nucleic Acids Res.* **24**:32–39.
5. Miklos, G.L.G. and Rubin, G.M. 1996. *Cell* **86**:521–529.
6. Mushegian, A.R. and Koonin, E.V. 1996. *Proc. Natl. Acad. Sci. USA* **93**:10268–10273.
7. Reich, J.G. and Sel'kov, E.E. 1981. *Energy Metabolism of the Cell.* Academic Press, New York.
8. Palsson, B.O., Joshi, A., and Ozturk, S.S. 1987. *Fed. Proc.* **46**:2485–2489.
9. Boiteux, A., Hess, B., and Sel'kov E.E. 1981. *Curr. Top. Cell. Reg.* **17**:171–203.
10. Schellenberger, W., Eschrich K., and Hoffman, B. 1981. *Adv. Enzyme Regul.* **19**:257–284.
11. Lauffenburger, D.A., and Linderman, J.J. 1993. *Receptors: Models for Binding, Trafficking, and Signaling.* Oxford University Press, New York.
12. McAdams, H.H. and Shapiro, L. 1993. *Science* **269**:651–656.
13. Tyson, J.J., Novak, B., Odell, G.M., Chen, K., and Thron, C.D. 1995. *Trends Biochem. Sci.* **21**:89–96.
14. Lee, I.-D. and Palsson, B.O. 1991. *Biomed. Biochim. Acta* **49**:771–789.
15. Fell, D. 1996. *Understanding the Control of Metabolism.* Portland Press, London.
16. Varma, A. and Palsson, B.O. 1994. *Bio/Technology* **12**:994–998.