
Developmental and physiological circuits: dissecting complexity. A report on a talk given by Dr Leroy Hood

BY JOHN SKÅR¹ AND PETER V. COVENEY²

¹*Karolinska Institutet, 171 77 Stockholm, Sweden*

²*Centre for Computational Science, Department of Chemistry,
University College London, 20 Gordon Street,
London WC1H 0AJ, UK*

Published online 2 May 2003

Dr Leroy Hood, the President and Director of the Institute for Systems Biology in Seattle, WA, USA, also spoke at the Nobel Symposium on self-organization. The main points of his talk are summarized in this paper.

Keywords: complexity; systems biology; information science; interdisciplinary tools

1. Introduction

Understanding of developmental and physiological circuits has become a major concern in biological and medical research. One approach, pursued by Hood and reported here, is to systematically dissect complexities in biological contexts by making use of systems-biology thinking and techniques.

The term ‘systems biology’ is not a new one, but in the last few years its use has become increasingly familiar. This is largely thanks to the growth in availability of information that makes it possible for researchers to define all of the elements present in a particular system, interrogate the inter-relationships between the elements, and to view the system in the context of its interactions with other kinds of systems.

The general idea behind systems biology is basically very simple. Both experimental and theoretical biology are relevant when considering the concept of systems biology. Theoretical biology plays an important role in encouraging the development of new concepts in biology, based on real data and pertaining to real situations. The results of the Human Genome Project provide many opportunities to apply the concepts of theoretical biology.

2. The genome project

The genome, one of the most remarkable digital codes known, consists of a collection of 23 pairs of chromosomes. In humans these chromosome pairs are present in most types of cell, equating to approximately three billion nucleotides.

One contribution of 18 to a Theme ‘Self-organization: the quest for the origin and evolution of structure’.

The genome project acted as a catalyst for change in the science of biology. As a result of the project, biology has been transformed into, and can be viewed as, more of an informational science. The genome project also changed the way biology is practised, by highlighting the usefulness of studying model organisms whose genome, or digital source code, is known and can be understood by means of genomics as a tool for deciphering biological complexity. Theories of evolution state that we all descended from a common origin and thus all employ a fundamentally similar set of basic informational strategies. Hence, in the context of humans, it is possible to use yeast to look at very fundamental aspects, and to turn to mice to study more complicated higher developed aspects of physiology and development.

In addition, the genome project spurred on the development of high-throughput tools, such as the DNA sequencer, an instrument developed in 1986 by Hood and his colleagues, which uses four different fluorescent dyes to colour code the four letters of the DNA language and read them out in a simple way. These developments are discussed below.

3. Biology as an informational science

The most important attitude change resulting from the genome project is the development of the concept of biology as an informational science. The core digital information present in the genome is a digital code that is ultimately knowable. By working with this source code, biologists can draw on knowable, ascertainable fundamental core knowledge and thinking about how biology actually occurs to gain a real understanding of biological processes.

The digital source code includes two fundamental types of information: the information contained in genes and information about regulatory networks. Genes are the molecular machines that execute the properties of life. The human genome, the ‘toolbox’ of genes, includes around 40 000 different genes. To a first approximation, the genomes in all metazoa, or multicellular animals, are very similar.

Regulatory networks include two features: proteins known as transcription factors and the transcription-factor binding sites to which these proteins bind. The collection of the transcription factors, and more importantly, their organization and disposition via these binding sites specifies the nature of the gene. The organization of the transcription-factor binding sites is digital. As a result, the process of regulation, in turn, can be considered to be fundamentally digital, because it involves changing the organization of the transcription-binding sites.

Regulatory networks evolve much more rapidly and in a different way from genes. A comparison between a chimp and a human illustrates this point. The two species diverged roughly six million years ago and share a large proportion of their genes, yet the structure of their brains is remarkably different, suggesting that the regulatory networks that control the developmental programs for creating a brain have changed in quite a striking way in six million years. One strong possibility is that the differences were caused by a concerted reorganization of transcription-factor binding sites. Transcription-factor binding sites evolving in some giant kind of combinatorial game may also have been responsible for the spectacular explosion of life that occurred *ca.* 650 million years ago at the Precambrian–Cambrian boundary at the beginning of the Palaeozoic Era, when a spectacular number of different multicellular organisms of different shapes, sizes and contents appeared in a relatively short period of time.

(a) Information hierarchy in biology

Biological information is of two fundamental types: the core digital information in the regulatory networks, and information about the environment, defined as anything that is outside this digital core.

Biological information is organized in many different levels of hierarchy. The hierarchy is organized so that information flows from DNA to RNA, to proteins, to proteins interacting with one another, to sets of proteins interacting to carry out a particular phenotypic function in biomodules and then to networks of biomodules within the cells. In order to understand the system it is necessary to capture information from as many of these different levels as possible, and integrate it. Finding ways to capture this information, and even more importantly, finding methods to integrate it, display it graphically and, ultimately, to mathematically model it, is one of the major challenges in systems biology.

An enormous amount of additional information comes in from the environment throughout this information hierarchy. Environmental information falls into two categories: signals and environmental factors. The signals are deterministic, that is, once received the consequences are inevitable. The environmental factors are stochastic, that is they generate randomness. In sophisticated systems, like the immune system, this randomness can be turned into information that drives very powerful kinds of mechanisms.

Stochastic events interpose themselves at every level of information hierarchy. One of the profound challenges that must be faced as biology moves towards considering global datasets is finding ways to distinguish signals from noise. Gaining an understanding of the historical context of biological systems will also be crucial. Comparative genomics is an important tool for deciphering the evolution of regulatory networks. Studies of the comparative genomics of contemporary organisms will also make it possible to extrapolate back and gain an idea of the mechanisms by which organisms evolve.

4. The development of high-throughput tools

In terms of technology, miniaturization, parallelization, integration and automation are key goals in all areas of biology. The benefits provided by improvements in analytical tools are already becoming apparent. For example, the development of analytical tools that work at higher throughputs means that more information is becoming available more quickly. These developments in analytical tools are important because they will provide access to the digital information of all living organisms. This access to digital information is already helping to transform the study of biology.

The pace of instrumentation development has been impressive. For example, the throughput of the DNA sequencers used today is roughly 6000 times greater than that of the first sequencers developed in 1986. And in the next five to seven years, thanks to developments in nanotechnology, it may become possible to carry out sequencing of single DNA molecules. This would lead to at least another 6000-fold increase in throughput. This would mean that just seven years from now we would be able to sequence an entire human genome in a day and at a cost of perhaps less than US\$5000.

Another important technological development involves the use of DNA microarrays. These take advantage of either photolithography or inkjet technology, to

synthesize hundreds or thousands of fragments of DNA on a 1 cm² glass chip. By taking advantage of the molecular complementarity of DNA, patterns of gene expression can be studied. Using microarrays it is possible, for example, to look at a normal prostate cell and a cancer prostate cell and describe, in a semi-quantitative way, the differences between the way the genes behave in the two types of cell. Access to these types of global datasets will transform our understanding of biology. The access to high-throughput tools has also spurred on developments in new areas of study, such as proteomics and comparative genomics.

5. New areas of study

Proteomics, the systematic study of protein structures, is another new tool. Proteins, naturally occurring polymers made up of amino acids that largely determine the structure and function of cells, are quite different from DNA sequences, which encode them.

The study of proteins generates data in a diversity of forms, and also presents many technical challenges. Tools to make it possible to identify and make quantitative comparisons between all the proteins in a cell, to characterize modifications, and to study their interactions, are being developed. These developments may eventually be applied on a more global scale to study the nature of all protein interactions, protein DNA interactions and transcription factors.

Comparative genomics is another potentially powerful area of study. In the next five to ten years, techniques will be developed to make it possible to take any genome and be able to separate out and reassemble the fundamental types of digital information in genes and regulatory networks, in order to provide insights into the logic of life for specific organisms. By comparing the results of genomic studies on hundreds or even thousands of organisms, it will be possible to begin deciphering the subtleties of regulatory networks.

Several biological systems are already beginning to be studied in this way, and the results are revealing new information about the way organisms function that could never have been gained by studying one gene or one protein at a time. For the first time it is becoming possible to see some of the very detailed and intimate interactions that occur within cells, thus shedding light on the nature of the dynamic interactions that occur within cells.

Mathematics is also likely to play a key role in systems biology. Some researchers are already using the results of proteomic and genomic studies to help them develop mathematical formulations to try to predict the behaviour of genes. However, the type of mathematics relevant to many aspects of systems biology is not familiar to many biologists. This highlights the importance of interdisciplinary collaborations in systems biology.

6. Integration and interdisciplinarity

Because it encompasses so many types of data and takes in such a broad range of investigations, research in systems biology is difficult to carry out in the context of classical academic constraints. In order to further the development of systems biology, a number of important questions need to be considered. On the purely technical side there are questions about the best ways to combine biology, technology

and computation. For example, what is the best way to create the high-throughput platforms which are essential to systems biology? How can they be kept up to date? What are the best ways to capture different levels of biological information? What are the best ways to integrate them, graphically display them and model them?

There are also important issues to consider about the best ways to create a cross-disciplinary faculty that works well as a team and whose members communicate effectively with each other. And in terms of scientific practice and methodology, it is also important to consider how best to reconcile the types of global analyses carried out in systems biology with the conventional methods of hypothesis-driven science.

The solution to all these dilemmas, Hood suggests, is to ensure that there are leaders in each of these areas developing the tools for tomorrow, bearing in mind that in order to be useful any new developments must also be integrated into the high-throughput platforms, and that it is important to develop and maintain partnerships with industry.

7. Educating biologists

Moving towards a systems-biology approach and beginning to think about biology in terms of an informational science, Hood believes, will not only lead to new insights in biology. It also offers a wonderful opportunity to instigate changes in the way biologists are educated. For example, more emphasis might be put on mathematics. With huge datasets to handle, there will be some really fascinating computational problems to deal with in the future. The formulation of mathematical approaches that can predict the behaviour of a biological system relies on one being able to develop ways to extract information in real time and finding ways to integrate biological information about genomes, physiology and development with environmental information, and the mathematics needed for doing this is not the type of mathematics generally taught to biologists.

Adopting the systems-biology approach and broadening the biology curriculum will also offer a further advantage: it will help biology to attract the best students.