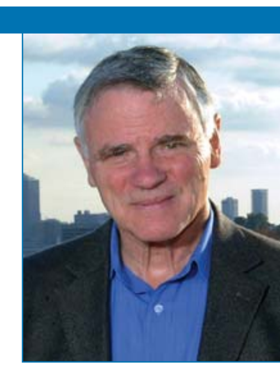
# Transforming Technologies for Biology and Medicine in the 21st Century



**Leroy Hood**
*Institute for Systems Biology, Seattle*

New directions in science are launched by new tools much more often than by new concepts. The effect of a concept-driven revolution is to explain old things in new ways. The effect of a tool-driven revolution is to discover new things that have to be explained. -Freeman J. Dyson

The biggest challenge for biology and medicine in the 21st century, as for most other scientific and engineering disciplines, is complexity. Biology is uniquely challenged and is strongly positioned to attack the problems of biological complexity with new strategies (a systems view of biology and medicine), an emerging integrative view of biology (as an informational science), and new technologies for measurement and visualization as well as the emergence of computational and mathematical tools for dealing with complex data sets. These include data acquisition, validation, storage, mining, integration, and descriptive, graphical, or mathematical modeling. I believe biology's ability to meet this challenge will make it a dominant science in the 21st century. But what exactly are the novel strategies and new technologies that are transforming biology?

Biology and medicine are moving steadily towards a systems' approach to understanding the fundamental features of life—holistic rather than atomistic—and embracing the view that information is the fundamental feature of life. This informational view has three aspects. First, there are two types of biological information—the digital information of each individual's genome and the environmental information that modifies this digital output. Understanding biology requires understanding how the digital genome and the environment signals are integrated leading to the five fundamental features of life: evolution, development, physiological responses, aging, and disease. From a technical viewpoint, a grand challenge in biology is to understand how digital and environmental information are integrated. Second, the information of life is managed by biological networks that acquire, integrate, and pass information to simple and complex molecular machines that, in turn, execute the functions of life. Understanding biology is about understanding how these dynamically changing networks handle information. The measurements required to delineate dynamically changing networks again constitute an enormous technical challenge (e.g. interactions of proteins and microRNAs with each other, with regulatory DNA sequences and other informational molecules measured in time). Third, biological information is hierarchical and multiscalar—DNA to RNA to proteins, to interactions, to networks, to tissues and organs to individuals to populations and to ecologies. To truly understand how complex biological systems function, one must integrate information from each of these relevant levels to understand the relative contributions of the genome and the environment at each stage. Another grand technical challenge of systems biology is to develop tools to make measurements and to visualize changes across this multiscalar informational space.

Systems biology began to emerge with the completion of the human genome project and the resulting complete parts list of genes, RNAs and proteins—thus enabling global or comprehensive measurements of the molecules of life (mRNAs and proteins) in response to biological perturbations—and the dynamic analysis and integration of this information leads to predictive models about how systems function. Systems biology requires the development of new technologies and computational tools to deal with data and its logical predictive models and hence requires a cross-disciplinary environment—bringing together biologists, chemists, computer scientists, engineers, physicists, and mathematicians. This cross-disciplinary environment is essential for carrying out the fundamental cycle of systems biology—namely, that biology drives or dictates the types of technologies that should be developed, and the data resulting from the technologies in turn drives the development of computational tools necessary for dealing with this data that ultimately leads to predictive mathematical tools of biological systems.

## Emerging segments of biology and medicine

New technologies have the potential to transform (e.g. the automated DNA sequencer that we developed in 1986 made the genome project possible, created the field of genomics, and enabled the field of proteomics). Several new emerging technologies are now in the process of transforming biology and medicine yet again.

• Next generation DNA. The throughput of DNA sequencing is increasing exponentially and will within the next eight to ten years move to the point that the genomes of individual humans can be determined quickly and inexpensively. Two features of sequencing are critical to this change: the ability to highly parallize the sequencing process—so that in the future perhaps a billion DNA sequences can be determined simultaneously—and the ability to carry out accurate sequence reads—all possibly on single strands of DNA (to simplify sample preparation and remove the PCR amplification steps).

• Measurement of protein levels in complex biological mixtures (e.g. blood). Microfluidic, nanotechnology, antibody, and mass spectrometry approaches are being developed to enable the quantitative measurement of thousands of proteins from complex mixtures such as blood. For example, monitor reaction measurement (MRM) mass spectrometry has the capacity to identify perhaps 1,000 peptides (e.g. proteins) at the femptomole level in several hours. The DEAL protein (microfluidic and nanotechnology) chip has a sensitivity of low femtomoles, a dynamic range of perhaps 106, and a sample analysis time of about 15 minutes and will in a few years be capable of measuring hundreds to thousands of proteins.

• One of the major limitations in modern biology is the ability to generate effective protein-capture agents such as antibodies. For example, the DEAL chip will require two good antibodies for each protein assay—and generating antibodies with appropriate specificity and sensitivity is expensive and time consuming. New technologies need to be developed for generating protein capture agents. We and others are exploring a new approach—click chemistry—that can chemically join two low-affinity ligands (peptides or aptomers) to generate a bivalent protein-capture agent that has the affinity of the two ligand multiplied together ($10^6 x 10^5 = 10^{11}$). These are stable reagents that can be synthesized rapidly and in large quantities.

• Single-cell analyses. It is clear that biology in the past has averaged the properties of large numbers of cells to do biochemistry or molecular biology. It is equally clear that the cells of a population generally exhibit quite diverse quantitative traits. To understand biology on a deep level, we must move to single cell analysis. Microfluidic approaches using soft materials (PDMF) allow one to make valves, pumps, and mixing chambers; and these coupled with immobilized cell-surface antibodies have allowed the separation of cell types and even the characterization of individual cells. These approaches enable the beginning characterizations of DNA, RNA and proteins from single cells or even pairs of interacting, but distinct cells.

• *In vitro* and *in vivo* imaging technologies. The ability to visualize molecular and structural information in cells and in organisms permits the integration of molecular information with higher level phenotypic observations. Nanoparticles coupled to specific protein-capture agents offer interesting opportunities to push these imaging frontiers.

## The P4 transformation of medicine

A systems view of medicine suggests that disease arises as a consequence of one or more disease-perturbed biological networks in the disease organ. The disease-perturbed networks modify the patterns of information they expression—and do so dynamically as the disease progresses. Altered patterns of information expression account for the pathophysiology of the disease and open interesting possibilities for the diagnosis, therapy and eventually prevention of disease. Thus, I believe that the systems approach to disease, coupled with new measurement and visualization technologies and the computational and mathematical tools for analyzing large data sets, will transform medicine from its currently largely reactive mode to one that is predictive, personalized, preventive, and participatory (P4 medicine).

Let me extrapolate perhaps ten years into the future to describe a possible scenario for P4 medicine and simultaneously speculate about the trajectories of the technologies described above. Predictive medicine will have three major components. First, the analyses of the genomes of individuals, the delineation of their maternal and paternal chromosomal haplotypes (chromosomal linkages), and their correlation with the health and disease of the corresponding individuals will increasingly lead to the ability to predict the future health history of the individual (e.g a 70% chance of prostate cancer by the time you are 50). Second, the measurement of perhaps 2,500 blood protein concentrations (perhaps with a DEAL-like technology) will allow the blood to become a window for the instantaneous assessment of health and/or disease. The identification of organ-specific proteins that are secreted into the blood allows one to sample the behavior of the biological networks in the chosen organ. For each individual, a blood fingerprint of organ-specific proteins will exhibit one set of protein concentrations for health and a distinct set for each disease state of that organ (because different diseases perturbed different sets of biological networks and hence their protein outputs). If the human has 50 organs of interest to assess health and disease and we need to assay 50 organ-specific proteins for each organ, then we would want to make 2,500 protein measurements from a fraction of a droplet of blood several times a year for each individual. Obviously we will need approaches like the DEAL chip to enable analyzing the bloods from millions of individual patients per year. The analyses of individual blood cells will likewise provide insights into the infectious-disease history of the individual—both past and current. Finally, *in vivo* molecular imaging techniques will begin to provide an assessment of the location, extent, and distribution of disease.

## Personalized medicine

Personalized medicine arises from the fact that each of us on average differs by six million nucleotides from our neighbors—and consequently are susceptible to differing combinations of diseases. The technologies described above will allow each individual to be examined in depth. Just consider that in ten years or so we will have billions of data points on each patient, and we must develop the information technology for healthcare to reduce this enormous data dimensionality.

Preventive medicine indicates that once we understand the dynamics of disease-perturbed networks, we can begin to design drugs to re-engineer them to behave in a more normal fashion. Moreover, if you have a 90% chance of lung cancer by the time you are 60 years, we may be able to design drugs to prevent your relevant networks from becoming disease-perturbed: preventive drugs.

Participatory medicine means that the individual patient, if appropriately educated, can take a more responsible role in choosing his/her own health trajectory because there will be much more information and understanding of what health is. The focus on the future will be wellness and more individual responsibility.

P4 medicine will lead to the digitalization of genomic medicine, and together with many other aspects of P4 medicine, will lead to a sharp reduction in cost. Just when this transition will occur will in part be dependent on the evolution of the above technologies and in part dependent upon how society accepts P4 medicine. It will also transform the business plans of virtually every sector of healthcare.

In sum—it is new strategies for doing biology and medicine (systems biology), a new view of biology (biology is an informational science), and emerging technologies that are driving this P4 transformation in medicine. Many of the described technologies are changing exponentially—following Moore's Law—and the slope of these changes will in part determine the rapidity with which this P4 medicine revolution. If Moore's Law applies to any existing technology, it takes only ten years to reduce today's capacity to a mere 10% of the future existing capacity. In general, this makes predicting the ten-year future extremely challenging. Nanotechnology and microfluidics, with their ability to miniaturize, facilitate automation and integration of complex measurement procedures, reduce cost per measurement, and execute real-time measurements, are going to play a profound role in many of the above described technologies. The fascinating question is what are the newly emerging future technologies whose shapes and potentials are not yet clear? It will be a fascinating next 10-20 years in biology and medicine.