Leroy Hood

A Personal Journey of Discovery: Developing Technology and Changing Biology

Lee Hood

Institute for Systems Biology, Seattle, Washington 98103;
email: lhood@systemsbiology.org

Annu. Rev. Anal. Chem. 2008. 1:1–43

First published online as a Review in Advance on February 27, 2008

The *Annual Review of Analytical Chemistry* is online at anchem.annualreviews.org

This article's doi:
10.1146/annurev.anchem.1.031207.113113

1936-1327/08/0719-0001$20.00

Key Words

biological instrumentation and strategies, automated DNA, protein sequencers and synthesizers, cross-disciplinary biology, systems biology, P4 medicine

Abstract

This autobiographical article describes my experiences in developing chemically based, biological technologies for deciphering biological information: DNA, RNA, proteins, interactions, and networks. The instruments developed include protein and DNA sequencers and synthesizers, as well as ink-jet technology for synthesizing DNA chips. Diverse new strategies for doing biology also arose from novel applications of these instruments. The functioning of these instruments can be integrated to generate powerful new approaches to cloning and characterizing genes from a small amount of protein sequence or to using gene sequences to synthesize peptide fragments so as to characterize various properties of the proteins. I also discuss the five paradigm changes in which I have participated: the development and integration of biological instrumentation; the human genome project; cross-disciplinary biology; systems biology; and predictive, personalized, preventive, and participatory (P4) medicine. Finally, I discuss the origins, the philosophy, some accomplishments, and the future trajectories of the Institute for Systems Biology.

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, Imagined Worlds

1. INTRODUCTION

I began to appreciate the beauty of chemistry in high school when, as a senior, I helped teach a sophomore biology class. I remember teaching the class using a 1956 *Scientific American* article on the structure of DNA. That article catalyzed the realization that the core of life was the DNA molecule with its fascinating chemistry of complementarity. Although I doubt I fully understood what that meant at the time, it was clear DNA was a beautiful molecule. This awakening pushed me toward the study of biology, but biology was always embedded in the context of chemistry. With this frame of reference grounded in chemisty, I was able to pioneer technology developments in biology and to grow up with the biotechnology industry. What follows is a personal overview of my career trajectories as they relate to biology, technology, paradigm changes, the creation of new companies, and the founding of organizations to change how biology is done. I discuss these topics chronologically as they emerged to provide the context for their origins.

I subscribe completely to Dyson's comment (see quotations above) (1) that new directions in science are launched by new tools much more often than by new concepts. Much of my career has been focused on developing instruments to decipher biological information, then applying them to fascinating biological problems—an endeavor made possible by a series of wonderful colleagues who were largely responsible for our success in pioneering new instruments and strategies for doing biology. Many of my colleagues (indicated in **Table 1**) went on to become leaders in technology development and application in both academia and industry.

Before continuing, let me say that in looking back it is quite easy to revise history in the context of our current understanding of the issues. Although I have tried to avoid this, I suspect it is impossible not to occasionally provide a compelling rationale for what were often intuitive decisions or decisions made on different grounds.

2. THE BEGINNINGS

I grew up in small towns in Montana where my parents always encouraged me to do well in school and gave me the freedom to explore many different dimensions of life. My father was an electrical engineer with the Mountain States Telephone Company and taught courses in electrical circuitry that I took while in high school. I was not interested in engineering, but these early courses probably provided me with a conceptual framework for my later thinking about systems biology. In high school I was encouraged to explore my potential by three outstanding teachers (in math, chemistry, and social studies and history), one of whom was instrumental in persuading me to attend the California Institute of Technology (Caltech).

Table 1 Colleagues in instrument and strategy development

Ruedi Aebersold	Protein blotting, protein microsequencing
Bruce Birren	Pulse-field gel electrophoresis
Alan Blanchard	Ink-jet DNA synthesizer
Richard Bonneau	Computational tools for protein folding
Ian Clark-Lewis	Long peptides
Cecilie Boysen	BAC shotgun sequencing
Nat Goodman	Database development and computational biology
Pat Griffin	Mass spectrometry and proteins
Mike Harrington	Two-dimensional gels and proteins
Suzanne Horvath	DNA synthesizer
Henry Huang	DNA sequencer (Maxam–Gilbert and Sanger approaches)
Mike Hunkapiller	Gas-liquid-phase protein sequencer, DNA synthesizer, DNA sequencer
Tim Hunkapiller	Computational tools, DNA sequencing
Daehee Hwang	Computational tools for systems biology
Trey Ideker	Computational tools for systems biology
Karen Jonscher	Mass spectrometry and proteins
Rob Kaiser	Labeling DNA
Steven Kent	Protein synthesizer, long peptides
Joan Kobori	Primer-directed sequencing
Eric Lai	Pulse-field gel electrophoresis
Ulf Landegren	OLA/SNP analyses
Steve Lasky	Ink-jet synthesizer
Chris Lausted	Ink-jet technology and surface plasmon resonance
Greg Mahairas	BAC-end sequence mapping
Debbie Nickerson	DNA polymorphism analyses
Jared Roach	Strategies for genomic analyses
Lee Rowen	Shotgun sequencing
Jack Silver	Radiolabeled microsequencing
Lloyd Smith	DNA sequencer (Sanger)
Mark Stolowitz	Protein attachment chemistry
Bingyun Sun	Prefractionation for mass spectrometry
Paul Tempst	Protein microsequencing
David Teplow	Protein chemistry
Mike Waterfield	Solid-phase protein sequencing
John Yates	Spectrometry and proteins
Hyuntae Yoo	Mass spectrometry and blood diagnostic

Abbreviations: BAC, bacterial artificial chromosome; OLA, oligonucleotide ligase assay; SNP, single nucleotide polymorphism.

The move from Shelby, Montana, to Pasadena, California, provided a striking culture shock, but after a year at Caltech I began to appreciate my exceptional classmates and the easy access to outstanding faculty. I had Richard Feynman for physics, Linus Pauling occasionally for chemistry, and George Beadle for biology, and from the beginning I appreciated the power of conceptually oriented teaching. Ray Owen, an immunologist, and James Bonner, a plant physiologist, also helped me to appreciate the marvelous beauty and complexity of biology. At Caltech my career expectations were raised, and I was provided with an excellent background in math, physics, and chemistry. By my senior year I had decided that I was primarily interested in human biology, so I went to Johns Hopkins Medical School in an accelerated three-year program and was immediately immersed in science (as well as medicine). I also found immunology, cancer biology, and diseases of the nervous system fascinating. Although all these areas became central to my later career, I found immunology particularly intriguing because of my readings, specifically a detailed topic paper associated with a fantastic microbiology course taught by Barry Wood, one of the early pioneers in infectious disease.

As my studies progressed I became convinced that the mechanisms of antibody diversity could be explored by characterizing the homogeneous blood immunoglobulin proteins derived from tumors of the antibody-producing cells (plasma cells) present in both mice and humans. In my search for a graduate school that would allow me to follow this direction, I ran into Bill Dreyer, who had recently moved to Caltech. Bill proposed a simple and supposedly noncompetitive "Saturday afternoon project" ideal for a new graduate student, namely that I sequence the homogeneous immunoglobulins that could be purified from the blood of mice with plasma cell tumors induced by interperitoneal injections of mineral oil. I accepted the proposal and became Bill's first graduate student. My first task was to learn the many aspects of protein chemistry, including protein purification, protein peptide mapping, and protein or Edman sequencing. The Edman chemistry that I learned in exhaustive detail would become the mainstay of our efforts to automate and increase the sensitivity of protein sequencing.

My conviction in conducting this research was that by determining the antibody sequences of the light and heavy immunoglobulin chains, one could reverse-translate back to gene sequences and begin to decipher the mysteries of the origins of antibody diversity. I immediately started to characterize immunoglobulin diversity, which catapulted me into the middle of one of the most exciting and rapidly moving periods of molecular immunology. With these amino acid sequences I was able to begin making fundamental hypotheses about antibody diversity (e.g., two genes encoded one antibody chain, diversity was encoded in many germ line genes, and recombinational and possibly somatic mutational events contributed significantly to antibody diversity) (2–5). I had now sensed the excitement of discovery and the satisfaction of personally formulating new conceptual ideas in biology—a heady opportunity for a second-year graduate student—and I was hooked forever on science as a career. At that time, Bill Dreyer also gave me two dicta that have guided my career: (*a*) Always practice biology at the leading edge; it is more fun there. (*b*) If you really want to transform a biological discipline, invent a new technology that permits you to explore new dimensions of data space.

3. THE NATIONAL INSTITUTES OF HEALTH YEARS: 1967–1970

After Caltech I wanted to do a postdoctoral fellowship in Europe, but I had a medical degree and Vietnam War–era policies dictated that all young doctors go either to Vietnam or into the Public Health Service. I chose the latter course and went to the Cancer Branch of the National Cancer Institute at the National Institutes of Health (NIH). I was given an independent position, established a first-rate protein chemistry lab, and continued working on molecular immunology. My tenure at the NIH resulted in two additional, unexpected benefits: First, I had the pleasure of meeting many young physicians who would become leaders in U.S. medicine. Second, I had time to think about what I would like to do with the rest of my career. I decided that I would like to create a laboratory where I spent half my time on molecular immunology (or other biologies) and half my time developing new technologies. After looking at several schools I realized that the Division of Biology at Caltech provided an ideal environment for this dual-research approach.

4. THE CALTECH RESEARCH YEARS: 1970–1992

At the outset of my Caltech career, I told the Chairman of Biology who hired me, Bob Sinsheimer, that I wanted to devote half my time to biology and half to technology. Based on my deliberations at NIH and my graduate experience with Bill Dreyer, my thinking about technology development was driven by two convictions: (*a*) The frontier needs of biology should determine which new technologies should be developed (and once developed, those technologies can revolutionize biology). (*b*) New technologies should focus on deciphering one of the several different types of molecular (or chemical) biological information (e.g., DNA, RNA, proteins). Initially, because of my expertise in protein chemistry, I thought about how to improve the Edman sequencing chemistry and its automation. By the mid-1970s I began to include protein synthesis, DNA synthesis, and DNA sequencing in my grand vision of biological instrumentation. "Decipher," in this case, meant determining the linear order of subunits in nucleic acid or protein digital strings so as to formulate hypotheses about the nature of gene structure, function, or evolution. It could also mean synthesizing smaller fragments of these DNA or protein strings so as to use clever new biological strategies (later in conjunction with recombinant DNA techniques) to further enable digital string characterizations.

Later I became interested in instrumentation for the quantification of mRNAs (ink-jet technology) and proteins (mass spectrometry). I wanted to develop instruments that would automate the chemistry of these processes. In doing so, their analytic throughput as well as the repetitive efficiencies of these synthetic or sequencing chemistries would be increased. This would in turn enable researchers either to synthesize or sequence longer strings (e.g., proteins and DNA), or to sequence regular strings with less starting material (e.g., proteins). For example, molecular immunologists in the 1970s and 1980s needed to be able to sequence small amounts of protein and, from such a protein sequence, clone the corresponding genes (see the following section). Likewise, immunologists needed to characterize the large and complex

gene families that encoded the antibody, major histocompatibility complex, and T cell receptors. These were the major biological drivers in my decision to develop the DNA and protein sequencers and synthesizers. Below, I describe each of the four instruments we developed during the Caltech years.

4.1. Protein Sequencer

At the time I moved to Caltech, my chemical expertise resided in the field of protein chemistry, so I logically began with the development of an automated protein sequencer with greater sensitivity and reliability. The challenge was twofold: (*a*) better efficiency of the cyclic Edman chemistry that cleaved amino acid residues one at a time from the N terminus of the protein and (*b*) the development of valves for the automated sequencing instrument that were leak proof and resistant to the corrosive Edman reagents and that prevented the reagent mixing that decreased the repetitive yield of the cyclic sequencing process and, hence, the length of chains that could be sequenced. The ultimate objective was to develop an instrument that was 100-fold (or higher) more sensitive than existing approaches for analyzing proteins. We played with a variety of approaches to improve the Edman sequencing process, including unsuccessful attempts involving solid-phase protein sequencing and radioactive protein sequencing (6). When Mike Hunkapiller joined the lab, progress in automated protein sequencing accelerated. Mike was an organic chemist who quickly gained engineering expertise, solving the valve challenge and getting a handle on the Edman chemistry. When we combined these efforts with Bill Dreyer's concept of a gas-liquid-phase protein sequencer (rather than the conventional spinning cup), we created an instrument that was approximately 100-fold more sensitive than its predecessors (7, 8).

The gas-liquid-phase protein sequencer, because of its increased sensitivity, allowed us to explore a series of new fields in biology because of the availability in low levels of many fascinating proteins that previously could not be sequenced. With the advent of recombinant DNA techniques, the sequencing of low-abundant proteins could lead to the cloning and characterization of the entire corresponding genes. We sequenced a human blood hormone, platelet-derived growth factor, and showed that its N-terminal sequence was nearly identical to that of an avian oncogene, *v-cis*. This observation generated for the first time the hypothesis that oncogenes are normal genes of human growth and development subject to control by a cancer virus (9). It was also the first time that a string search of a new protein sequence was carried out against a database of preexisting protein strings to learn interesting new biology—the beginning of bioinformatics.

With Stanley Prusiner, we sequenced the prion protein (10), which enabled the cloning of the gene (11). This eventually led to the hypothesis of an infectious protein—infectious because its misfolded structure encoded a catalytic ability to convert normal prion proteins to the infectious form, thus generating an autocatalytic disease that led to neural degeneration. Stanley Prusiner won a Nobel Prize in 1996 for this work. Prion disease later became our flagship initial study of a systems approach to disease (see below).

We also sequenced erythropoietin, providing useful information for Amgen in its eventual cloning of the gene and the creation of biotechnology's first billion-dollar

drug. Our sequencing of the four chains of the torpedo acetylcholine receptor (12) led to the now-classic work of Professor Nomura at Kyoto University, who cloned the genes encoding this receptor and many other neuroreceptors, thereby transforming our understanding of important aspects of how the brain functions.

We were the first to sequence the α and β interferon proteins (13, 14), facilitating the cloning of their genes as well as their eventual deployment as useful drugs for cancer and multiple sclerosis. We were also the first to sequence a hematopoietic colony-stimulating factor, which led to cloning and its application as a useful drug.

In each case these sequences initiated new approaches to biology, and in many cases they generated interesting biotech drugs. These are but a few examples of the many other fascinating proteins that we microsequenced during the late 1970s and 1980s (15–18).

Below I describe the steps we took to commercialize the protein sequencer, which resulted in the creation of an extraordinarily successful company, Applied Biosystems, Inc. (ABI). The productive partnership between the Hood lab at Caltech and ABI led to the development of robust instruments for each of the three technologies described below.

4.2. DNA Synthesizer

In the late 1970s, Marvin Caruthers at the University of Colorado had just developed the phosphoramidite chemistry for DNA synthesis. I proposed that he teach my technician, Suzanne Horvath, how to do manual DNA synthesis so that we, in turn, could automate this process. Marvin argued that he could easily teach the procedure in a week to anyone who needed to synthesize DNA, and because he felt the demand for oligonucleotides was never going to be significant, why bother with automation? I convinced him otherwise. Suzanne learned the procedure and then, together with Mike Hunkapiller, designed an instrument for the repetitive DNA synthetic procedure (19). The challenge, once again, was designing valves that could withstand corrosive chemicals and maintain the separation of reagents from various cycle steps to improve the repetitive yield. Mike's earlier experience with the protein sequencer valves was useful in this regard. Caltech put together a prototype, and ABI moved quickly to develop robust DNA synthesis instrumentation.

The DNA synthesis platform provided critical oligonucleotides for many aspects of the emerging recombinant DNA technologies. We quickly suggested a new sequencing strategy that employed oligonucleotide primers for primer-directed DNA sequencing (20). We saw that genes could be synthesized by joining overlapping oligonucleotides in a sequential and hierarchical manner. We cloned interesting genes by synthesizing degenerate oligonucleotides, reverse-translated directly from the protein sequence analysis of low-abundance proteins (see, e.g., Reference 11). It was obvious that the ability to synthesize DNA primers enabled the conceptualization and development of the DNA amplification procedure—the polymerase chain reaction—because this procedure required pairs of oligonucleotides. So once again a new instrument provided many new possibilities for biology.

4.3. Peptide Synthesizer

Steven Kent came to my laboratory in 1983 with a wealth of protein synthesis experience from the Merrifield lab at Rockefeller University. Kent was attracted to the idea that the repetitive chemistry of protein synthesis could be improved through superior instrumentation to the point that modest-sized proteins could be synthesized with high yields, thus both improving the chemistry and developing another instrument utilizing leak-proof and corrosion-resistant valves. Steve worked in the Hood lab together with ABI to develop this instrumentation (21). Once again, many of his challenges were similar to those arising from the automation of protein sequencing and DNA synthesis. The chemistry in each case was, of course, unique, and each improvement presented special chemical challenges. I remember a time when an ABI executive came to me with a field representative's assessment that there was a market for only a few peptide-synthesis machines per year (fewer than ten). Fortunately, my contrary convictions overrode this erroneous prediction (ABI sold approximately 70 in the first year), and development of the peptide synthesizer proceeded.

Steve and ABI developed a superb peptide synthesizer with high repetitive yields and hence the ability to synthesize long peptides or even small proteins. Perhaps the most spectacular result arose from Steve's collaboration with Merck to synthesize the HIV protease (99 residues). They did this so effectively that after purification, it was possible to crystallize the chemically synthesized protein and solve the crystal structure to a resolution of a few angstroms (22). From these data, Merck developed its antiprotease drug, which came to be one of the most successful AIDS drugs. Steve went on to synthesize several interleukins, as well as a variety of other proteins, and carried out a fascinating series of structure-function studies (23, 24), and we did some interesting peptide/DNA binding studies with zinc fingers (25).

4.4. DNA Sequencer

Certainly the most difficult Caltech instrument to develop was the automated DNA sequencer. In 1975, Maxam & Gilbert (26) and Sanger (27) developed the chemical and enzymatic (di-deoxy) manual approaches to DNA sequencing, respectively. In the late 1970s, Henry Huang, a postdoctoral fellow in my lab, began working to automate first the chemical and later the enzymatic approach. Henry was a biologist with some interest in engineering, but it gradually became clear that better knowledge of engineering and a far more sophisticated chemical expertise were required to complete this project. In 1982, I assembled a team including a chemist/laser expert (Lloyd Smith), an engineer/chemist (Mike Hunkapiller), a biologist with knowledge of computer science (Tim Hunkapiller), and myself, a molecular biologist—and one spring day we had a transforming conversation. Four central ideas emerged about a new proposed approach to automated DNA sequencing: (*a*) The DNA fragments of the Sanger reactions could probably be separated nicely by capillary gel electrophoresis. (*b*) The fragments could be labeled with one of four different fluorescent dyes, according to which base terminated the fragment. (*c*) All four colored bases could be detected together in a single capillary channel (the manual sequencing approaches

used radioactive reporter groups that required four separate lanes, one for each DNA fragment ending in a distinct base), thus standardizing the DNA fragment comparisons and increasing the efficiency of the sequencing process. (*d*) The four distinct classes of fluorescence-labeled DNA fragments could be distinguished by laser detection of the dyes, and this four-parameter dye space signal could be converted into DNA sequence by computational algorithms. Although these simple ideas emerged in a single afternoon, it took another three years before the practical details were solved and a prototype, capillary-based automated DNA sequencing instrument was developed (28). We had to develop a chemistry for coupling the dyes to DNA (29), identify four good dyes, design laser instrumentation for reading dye space, generate algorithms for converting dye space into DNA sequence, optimize the enzymology of polymerases for the extension sequencing reactions, and solve a host of other chemical and engineering challenges. Lloyd Smith played a central role in solving many of these problems. This team effort illustrated the power of a cross-disciplinary approach and the need for team science in solving a challenging technical problem (two points I return to below). During the latter stages of this effort, Mike Hunkapiller moved to ABI, which was very much a partner in developing the automated DNA sequencer. ABI took the lead in pioneering the robust instrument necessary for its commercialization.

The DNA sequencer truly changed biology by making the human genome project possible. The DNA sequencer also pointed the way toward the development of even higher throughput DNA sequencing instrumentation for the generation of the massive amounts of data essential for eventually finishing the human genome project. The genome project, through its genetic parts list, eventually paved the way for systems biology. It also made possible the current analyses of thousands of genome sequences from microbes, plants, animals, and even multiple humans; these, in turn, have transformed and are transforming many different fields of biology and medicine (30–32).

4.5. Integrated Microchemical Facility

In the early 1980s, my lab suggested that developing and coordinating or integrating the respective functions of these four automated instruments (DNA and protein sequencers and synthesizers) would lead to the creation of an integrated microchemical facility with a powerful capacity for moving from genes to proteins and vice versa (33). For example, if the protein sequencer was used to determine the amino acid sequence of an interesting protein, this protein sequence could then be translated via the genetic code dictionary into a degenerate oligonucleotide sequence; degenerate oligonucleotides could then be generated by the DNA synthesizer, and these DNA probes could then be employed with recombinant DNA techniques to clone the corresponding genomic or cDNA clones that could then be sequenced by the automated DNA sequencer. In a similar vein, genes could be translated into an amino acid sequence that could then be synthesized by the protein synthesizer, and the resulting peptide fragments could be used as antigens to generate specific antibodies for protein localization and characterization. Likewise, multiple degenerate oligonucleotide

probes could be used to clone large gene families. Thus, the integrated microchemical facility could be used to coordinate the chemistries of these four instruments in conjunction with appropriate recombinant DNA and biological techniques.

Interestingly, publication of the *Nature* paper that described this integrated microchemical facility (33) was delayed for almost two years because of reviewers' skepticism about several of the predicted integrated strategies. In fact, microchemical facilities became common throughout the academic and industrial worlds. The automation and chemical optimization of these sequencing and synthesis procedures also led naturally to the emergence of other high-throughput technologies, which arose either by parallelization (from 1 sequencing capillary tube to 96) or by more rapid serial procedures (often from miniaturization of reaction procedures or more efficient chemistries), thus providing a push for the eventual development of both genomic and proteomic high-throughput platforms capable of generating the large data sets required by systems biology. Below, I discuss our development of one of the most powerful high-throughput technologies in biology today: the ink-jet DNA synthesis technology that enables the synthesis of DNA chips. The microchemical facility became an early embodiment of the Caltech vision for transforming biology through automated instrumentation.

4.6. Commercialization of the Four Instruments

By the late 1970s we had developed the protein sequencer, were working on automated DNA sequencing and synthesis, and were thinking about peptide synthesis. A friend suggested that we think about commercializing these instruments to make them available to others. Caltech's president, Murph Goldberger, was skeptical: He argued that the role of an academic institution was scholarship and education, not commercialization. "But," he said, "you can try to commercialize if you wish." I went to 19 different instrument companies over the next year and all said no. I visited Beckman, Inc., three times before they told me not to come back. Then Bill Bowes, a venture capitalist from San Francisco, called me and said he knew of my failed attempts with the instrumentation companies, and he offered to provide $2 million to start a venture-backed company. Murph, once again, was cautious; he was reluctant to accept venture money because Harvard was then going through a messy debate about the venture-backed creation of a company called Genetics Institute. In the end, however, Murph relented, and just as we were about to sign an agreement with Bill Bowes, I gave a talk to the Caltech trustees on the vision of how these four instruments would change biology. Arnold Beckman, founder of Beckman, Inc., and a Caltech trustee, immediately approached me afterward and said that this new instrumentation was just what Beckman, Inc., needed. An awkward moment followed, and I said that his company had not appeared interested. A complex and difficult series of misunderstandings ensued between Arnold and Caltech (Beckman was a large donor to Caltech), but in the end, Bill Bowes and I started ABI—and it grew to be the most successful biotech instrumentation company worldwide. Indeed, the protein sequencer had been so well engineered that it was successfully produced in the second quarter of the first year of ABI's existence, and the company was in the black by the end of that year. All

turned out well for Caltech as well, because Arnold Beckman later donated more than $100 million to create the now well-known Beckman Institute at Caltech.

This experience taught me three important lessons. First, always discuss visions with the highest-level administrator or leader (e.g., CEOs, founders)—for they, if anybody, will be able to comprehend the vision's future potential and relevance to the company or organization. Second, it was fortunate that none of the extant companies accepted my offer of the instruments, as none would have been able to acquire the scientific talent, provide the sufficient resources, or focus entirely on the problem of developing these four instruments. ABI did a superb job in all these regards and represented the new organizational structure needed to realize this new instrumentation vision or paradigm change of developing and commercializing automated and integrated instrumentation—a vision that later encompassed high-throughput data-production instruments. This was the first of several paradigm changes in which I participated, creating new organizational structures when they were needed to catalyze the paradigm change. I went on to play a founding or cofounding role in creating more than 13 additional biotech companies (including Amgen, Systemix, Darwin Molecular, and Rosetta Inpharmatics) and thus effectively transferred academic knowledge to society. Third, I realized that there are five stages to the development of new instruments: (*a*) conceptualization; (*b*) the development of a prototype instrument to provide proof of principle; (*c*) the development of robust instruments that any biologist could use; (*d*) the conversion of the robust instrument into a high-throughput platform; and (*e*) the development of a completely novel and better approach for the relevant chemical instrument (where better could mean faster, cheaper, miniaturized, more highly parallelized, more effectively automated, and/or integrated with other chemistries or procedures), followed by repetition of the first four stages.

I came to realize that academia is superb at conceptualization and prototype development, but generally has neither the resources nor the systems engineering skills for robust commercial development and high-throughput conversion. Novel approaches to the instruments in development will likely arise once again in academia. I can illustrate this point by noting that conceptualization of the automated DNA sequencer was very inexpensive; prototype development probably cost approximately $500,000 (and I failed in two attempts to get this funded by NIH); commercial development cost approximately $75 million; high-throughput conversion probably cost several hundred million dollars altogether; and several new approaches to very-high-throughput DNA sequencing instrumentation are now under way. Conceptually, these new approaches came initially from academic labs. Thus, the creation of ABI was the critical step in realizing the potential of our instruments, and their widespread use throughout the scientific community resulted from our commercialization.

4.7. New Strategies Employing the Four Instruments

Not only did the four commercialized instruments open up new areas of biology when used directly as they were designed, they also enabled the pioneering of new strategies for carrying out biology, such as primer-directed DNA sequencing, the synthesis and assembly of long DNA strings including genes, the use of degenerate

oligonucleotides reverse-translated from protein sequences to clone genes, and the production of peptide antibodies (all discussed above). In addition, we developed a series of additional genomic and recombinant DNA strategies (34–42). At the protein level, Mike Hunkapiller, Ruedi Aebersold, Steven Kent, Mike Harrington, and Paul Tempst developed new approaches for the purification of proteins and their microsequence analyses (43–53). John Yates and Pat Griffin pioneered the use of the mass spectrometer as a tool for protein and peptide identification and characterization. Yeats also developed some of the earliest computational proteomics techniques (54–57). Also, we developed a pulse-field instrument for the size measurement of very large DNA fragments (58, 59). The significance of these developments is that our new instrumentation made it possible to create many new strategies for generating data at both the DNA and protein levels. Two of these strategies deserve special mention because they initially faced considerable skepticism from the relevant scientific communities as to their utility, but eventually played important roles in genomic analyses: the oligonucleotide ligation assay and the use of mate pairs (e.g., sequence data from both ends of large insert clones) to facilitate the mapping and sequencing of entire genomes.

When Ulf Landegren was a postdoctoral fellow in my lab, he proposed that enzymatic oligonucleotide ligation could be used as a means for identifying single nucleotide polymorphisms (SNPs) (60). His idea was that if two adjacent oligonucleotides were complementary to a target DNA, then they could be covalently joined at their abutting ends by the enzyme DNA ligase. If, however, a polymorphism were present at the 5′ end of the 3′ oligonucleotide, a mismatch in complementarity would prevent the ligase from joining the two oligonucleotides. Hence, one could synthesize two 3′ oligonucleotides, each with a distinct 5′ base complementary to one of the two SNP variants. Accordingly, if each of the two 3′ oligonucleotides were labeled at its 3′ end with a different fluorescent dye, then one could identify the SNP variants by the color of the 3′ probe ligated to its 5′ counterpart (a heterozygous individual with both SNP variants would ligate both colored 3′ ends, and a homozygous individual with just one variant would ligate the complementary 3′ end and exhibit just a single color). Debbie Nickerson, a visiting professor in my lab, later applied the oligonucleotide ligation assay in a variety of ways for effective and revealing SNP analyses (61). This procedure has been the basis for a variety of very-high-throughput SNP typing procedures, including one first developed by the company Illumina, which used it to make major contributions to the large-scale SNP mapping required to delineate the human haplotype SNP map (ascertaining the SNP linkage relationships on the maternal and paternal chromosome sets from individual humans).

The use of paired ends, or mate-pair information, to facilitate the assembly of genomic regions from collections of shotgun sequencing reads was initially proposed by Jared Roach, a graduate student in my lab (62). In shotgun sequencing, a genome, or a large insert clone derived from a genome, is randomly fragmented into short pieces, which are then subcloned into a suitable cloning vector and propagated as DNA templates for sequencing; from these, one obtains sequence reads of several hundred bases in length. These reads are then assembled by aligning and conjoining overlapping sequence strings, with the goal of reconstructing the original genome or large insert clone sequence. This task is facilitated by obtaining redundant coverage

of the genome from the shotgun reads, and is hindered by the presence of duplicated regions (repeats) or by regions with no sequence read coverage. My colleague Lee Rowen employed this approach in a classic sequence analysis of the 650-kb-long human T cell receptor locus (63)—by far the longest segment of DNA sequenced at the time—and these efforts were followed in our lab by the sequence analysis of the >1-Mb human (64) and mouse α T cell receptor loci and the major histocompatibility complex locus (65). We also did a fascinating species comparison of the T cell receptor loci (66). In the early days of genome sequencing, a single-stranded virus (M13) was used as the cloning vector because of its ease of propagation and purification. Roach rightly pointed out that if a double-stranded vector, such as a plasmid, were used for the propagation of DNA templates, then a sequence could be obtained from both ends of the genomic fragment, rather than obtaining only the single read provided with the M13 vector. The advantage to using a double-stranded vector lay in the ability of these mate pairs to increase the accuracy of the assembly process by localizing repeats (if one half of the pair aligned to a unique region) and gaps (if the mate pairs spanned a gap in the assembled sequence, then one could infer that these stretches of sequence were in proximity in the genome). Thus, the mate pairs allowed for the ordering and orienting of reconstructed blocks of genomic sequence assembled from shotgun reads.

The next leap for using mate-pair information came about with the development in Mel Simon's lab at Caltech of a cloning vector called bacterial artificial chromosomes (BACs), which can propagate large inserts up to ~250 kb (67). Cecilie Boysen, a graduate student in my lab, was the first to demonstrate that inserts propagated in BACs could be successfully sequenced and assembled (68). This success led Craig Venter and me to propose using BACs and BAC-end mate-pair sequences as reagents for mapping genomes in a way that would facilitate high-throughput sequencing of large chromosomes (69). The basic idea was the following: We can begin with the assumption that 30,000 randomly generated BAC insert clones, each approximately 200 kb in length, could be sequenced for 500 base pairs at either end. If the inserts of the BAC clones were randomly generated, then there would be a 500-base-pair sequence tag, on average, every 50,000 base pairs across the genome. Once a 200-kb BAC was completely sequenced, on average, it would overlap with the end sequences of four other BAC clones—and one could simultaneously sequence out in both directions from tens or hundreds of nucleating BAC clones to generate dual growing points of overlapping sequences, sequentially allowing individual chromosomes to be sequenced. Although this approach was initially unpopular, we were able to obtain funding to sequence the ends of hundreds of thousands of human BACs (reduced to practice by Greg Mahairas in my lab), and thereby create what proved to be an invaluable resource for completing the human genome (70). In spite of some technical limitations to this strategy, the BAC-end sequencing approach has been and is widely used in whole-genome sequencing.

4.8. Systems Biology: Beginning Thoughts and Conceptualization

My labs at Caltech were located next to those of Max Delbruck, a Nobel Prize–winning physicist who pioneered the field of quantitative biology. Max was never very

impressed with immunology. He was skeptical about whether the deep problems of immunology—the immune response, tolerance, and autoimmunity—could ever be solved by a one-gene- or one-protein-at-a-time approach. I argued that we had done very well in coming to understand the mechanisms of antibody diversity, and he replied that diversity was an easy problem and not a deep problem of immunology. I gradually concluded that Max was correct at least in part, and I started thinking about what it would take to successfully attack these big problems in immunology (which remain largely unsolved today). I decided that one needed to study the systems of genes or proteins that mediate immune phenotypes, and not just individual genes or their protein counterparts. But we did not have the tools for identifying the components of biological systems and their interactions in the mid-1980s, nor did we have a complete parts list of all genes nor, by inference, all proteins. We needed to be able to do far more comprehensive analyses of the behaviors (changes in structure, expression levels, interactions, and cellular localizations) of mRNAs and proteins (and this also led in part to the conceptualization of the ink-jet DNA array technology, discussed below).

This systems thinking also made the idea of the genome project attractive because one could more or less completely define a parts list of genes and, by inference, mRNAs and proteins—and thus hope to carry out comprehensive or global analyses (see below). I submitted a couple of systems-like biology grants to the National Science Foundation (NSF) that did not do very well because the reviewers were unclear as to what I was proposing (it was not clear that I completely understood what I was proposing either). In 1991 I wrote a chapter in the book *The Code of Codes* (71) (which was about the human genome project and was coedited by Dan Kevles and me) that contained a description of systems biology that could be used today (although I did not start using the term systems biology until a few years later). Our current view of systems biology emerged slowly over the late 1980s and early 1990s, fueled by the creation of an NSF Science and Technology Center (STC) at Caltech in 1989 and then further matured by the establishment of the cross-disciplinary Department of Molecular Biotechnology (MBT) at the University of Washington (see below).

4.9. The Human Genome Project

The development of the DNA sequencer presented me with two interesting new opportunities to change how biology is done: (*a*) the genome project and (*b*) cross-disciplinary biology. The first opportunity arose when, because of our development of automated DNA sequencing, Bob Sinsheimer, now the Chancellor of the University of California at Santa Cruz, invited me to the first-ever meeting on the human genome project in the spring of 1985. Bob was considering setting up an institute at Santa Cruz to sequence the human genome, and he had invited 12 experts to discuss this possibility. Over several days the group came to two conclusions: first, that the human genome sequence was technically feasible, although difficult (the prototype DNA sequencer had just been developed), and second, that the group was evenly split as to the advisability of carrying out this project. I came away from this meeting with several impressions: (*a*) The genome project would certainly drive the development

of DNA sequencing and other technologies; (*b*) it would require the development of an array of new computational tools; and (*c*) it would provide a complete list of all human genes, a necessity for carrying out some of the systems approaches to biology I was just starting to consider (discussed in detail below).

In 1985, Wally Gilbert, Charlie Cantor, and I, along with others, began talking to the community of biologists about the genome project. I was surprised to find that perhaps 90% of the biologists bitterly opposed it on several grounds, namely that it was big science and that big science was inherently bad and would take the funding from hypothesis-driven small science. Moreover, the opponents argued, nothing interesting would come from the genome project; therefore, it would be impossible to recruit talented scientists to work on it. In my view, neither argument had any merit. NIH was initially firmly opposed to the project, and, indeed, it was the Department of Energy that really championed this proposal in its early days. A National Academy of Sciences committee, comprising both opponents and champions of the project, unanimously endorsed it after a year of deliberation. This endorsement brought NIH to the genome table, and an essential component of NIH's subsequent success with the genome project was the establishment of a brand-new institute at NIH to oversee this process. The proposed 15-year project began in 1990 and was more or less complete in 2004.

The striking lesson I learned from this experience was how conservative most scientists are, and how difficult it can be to reason with those who operate primarily from the biases of their past experiences, rather than thinking about the potential of future possibilities. In retrospect, another interesting point emerged. A friend told me in 1985 that it would take 100 years to sequence the human genome; with the technology we had at that time, he was probably correct. The real driver of change in biology (or science in general) is technology (see the quotation by Dyson at the beginning of this article), and what most people fail to understand is that some technologies can change exponentially at certain periods of their development. Thus, our ability to predict the future of a field depends very much on understanding how rapidly the driver technologies for a field are changing. For DNA sequencing technology, the period from 1986 to approximately 1998 was one of those periods of exponential change. Hence, if one could gauge the nature of this change, one could make predictions about future possibilities that would appear excessive to those who do not understand the dynamics of exponentially changing relevant technologies. One of my strengths was that I often saw which were the important technologies (and important ideas in biology) and how these technologies were changing or were going to change. Interestingly, today we are going through a similar period of exponential change in the increasing throughput of DNA sequencing, which will transform predictive medicine (as I discuss below).

We played an important role in learning how to sequence large DNA fragments (62–66): We were the first to do an evolutionary analysis of a 100-kb fragment from an important region in the human and mouse α/δT cell receptor loci and indeed defined most of the parameters used today in cross-species comparisons (72). We then established one of the 16 centers to sequence the human genome and sequenced significant portions of chromosomes 14 (73) and 15 (74), which contributed to the

complete human genome sequence (75). Thus we developed the instrumentation and pioneered its applications to biology.

4.10. Cross-Disciplinary Biology

The second insight I had from developing the DNA sequencer was the power of bringing together scientists from different fields to attack a technically challenging problem. I applied this insight in the context of my lab and as a result it grew to be very large (as we were doing both biology of several different flavors and broad-spectrum technology development). In 1987, I decided to compete for an NSF STC—an idea pioneered by Dick Zare, a first-class chemist at Stanford. The idea was to integrate science (biology in my case) with the development of the appropriate technology. In addition, the STC strategy emphasized industrial strategic partnerships and facilitating K–12 science education. This STC grant was the most effective grant I ever received because of its flexibility and the breadth and relevance of its requirements. This grant, together with my experience in developing instrumentation, continued to transform my thinking about how to do science. I realized the importance of bringing to biology a cross-disciplinary environment wherein biologists, chemists, computer scientists, engineers, mathematicians, and physicists could focus on using biological challenges to drive the development of relevant technologies and computational or mathematical tools. In such a cross-disciplinary environment, it is imperative for each scientist to learn to speak the languages of the other scientists and engineers, and work together in teams to attack hard biological problems in an integrated manner. I approached a different Caltech president, Tom Everhart, to argue that Caltech should start a new cross-disciplinary biology department with a cross-disciplinary faculty focused on biology. Tom said that it sounded like a good idea, but that I would have to persuade my colleagues. The chemists and engineers thought it was a good idea, but the biologists opposed it for reasons that were never completely clear to me.

Let me stress that not all Caltech biologists were opposed or indifferent to my interest in technology. Eric Davidson, a long-time loyal friend and wonderful collaborator on the genomics and systems biology of the sea urchin (76, 77), was a strong supporter of my various efforts throughout my 22 years at Caltech. Eric always appreciated and readily adapted and employed new technologies, and he continues to do so today. Just last year, he led a magnificent effort to sequence and analyze in enormous detail the sea urchin genome (see the Nov. 10, 2006, issue of *Science*). Over the past 35 years, Eric has fashioned the sea urchin into a superb model for development in metazoan organisms by dint of brilliant experimentation, always applying cutting-edge technologies. Sea urchin development today, as pioneered by Eric, is one of the outstanding examples of a systems approach to biological complexity.

However, the opposition of the leadership in biology to a possible cross-disciplinary department at Caltech presented me with the most difficult professional decision I ever had to make. Should I stay at Caltech with great students and colleagues, or should I move elsewhere to create a new cross-disciplinary department? I decided to follow my cross-disciplinary convictions. After looking at several schools, the University of Washington (UW) became a possibility, partly because a former

student—Roger Perlmutter, then Chairman of Immunology at the medical school—pushed the idea of my recruitment. I interviewed with the dean of the school of medicine, Phil Fialkow, and at the end of our interview he said that a cross-disciplinary biology department would be far too complex for a medical school. I went home disappointed, but Phil called me a few days later and said that he had made a mistake and that he wanted to fly to Pasadena and talk to me about rectifying it (this was a remarkable event—a dean admitting he had made a mistake—I was really impressed).

We met and came to an agreement that I would consider a move to the UW. I would first build the cross-disciplinary department (MBT) and then, after four or five years, I would receive additional space and could expand to build up systems biology (this latter agreement was informal and never documented). Phil arranged for me to give three John Dantz lectures at the UW on the future of biology—all were attended by Bill Gates. After the last lecture, Bill and I had a fascinating dinner that lasted about three hours; we talked at length about our views of science and engineering. Bill then offered $12 million to support this new department. So in 1992, after a wonderful 22 years at Caltech (in addition to eight more as an undergraduate and graduate student), I moved to the UW to found MBT.

5. THE MOLECULAR BIOTECHNOLOGY YEARS: 1992–2000

5.1. Getting Started

Moving to the UW was difficult: Not only was the cultural and scientific environment quite different from Caltech, but our space was not going to be ready for several years. The department was crammed into a modest space in the Washington Technology Center, staffed by a diffuse collection of many different types of scientists and engineers—some of whom focused more on commercialization than on academic science. Fortunately, many of my key colleagues moved with me from Caltech, some to be faculty members. We recruited wonderful additional faculty from a variety of institutions with many different areas of expertise, including genomics, computational biology, protein chemistry, and later proteomics, biology, computer science, and mathematics. At MBT my lab developed its fifth instrument: the ink-jet DNA arrayer.

5.2. Ink-Jet DNA Array Technology

Alan Blanchard was a computational biologist who came with me to MBT from Caltech. At Caltech he had started thinking about using common ink-jet printer technology for synthesizing arrays of oligonucleotides on glass or silicon chips (78, 79). The biological impetus was the desire to quantitatively analyze the concentrations of potentially all the mRNAs of humans or other organisms—one of the global technologies that is vital to systems biology. Prior to our efforts, Steve Fodor began developing a photolithography approach to DNA array synthesis for the company Affymetrix. The ink-jet project called for sophisticated surface chemistry: It required struggling with DNA synthesis chemistry in the context of the ink-jet printer environment (oxygen was lethal to the synthesis reactions) and the ability to computationally

drive and engineer the ink-jet printer. Alan persevered, learning chemistry and employing his knowledge of computer science, and ultimately produced a functioning prototype instrument. However, we were a long way from an effective and robust instrument, and commercialization to achieve these goals was essential. Ultimately we started a company, Rosetta, that continued developing this technology for commercial application. Rosetta eventually licensed it to Agilent, which made further significant improvements. One advantage of the ink-jet DNA arrays over their photolithography counterparts is that the efficiency of the synthesis process permits long oligonucleotides to be generated (60–70 mers versus 25 mers), and these have significant advantages for accurate hybridization and special applications. For example, the creation of ink-jet DNA arrays allowed for striking new opportunities in synthetic biology for the assembly of extremely long DNA fragments—the ability to synthesize everything from genes to gene families to genomes. The ink-jet approach is the core technology Agilent uses today for DNA chip synthesis. In my lab, Chris Lausted and Steve Lasky developed a less sophisticated but quite functional ink-jet synthesizer instrument (80) whose design was copied by several other institutions.

I stress another important point about high-throughput instrumentation: If one can sequence, synthesize, measure, or localize in a high-throughput mode, one will transform the nature of the biology that can be done. We can see this in the ways automated DNA sequencing and DNA arrays have changed evolutionary biology, developmental biology, genetics, and physiology over the past 15 years. Indeed, as I discuss below, in perhaps ten years we will be able to sequence individual human genomes rapidly and inexpensively; this will be an important element in transforming medicine.

5.3. The Struggle to Realize the Systems Biology Vision

In 1996—at about the time I was beginning to think of developing systems biology—Phil Fialkow tragically died in a blizzard in the Himalayas. Phil was a wonderful human geneticist who fostered a marvelous environment for the development of basic science as dean of the University of Washington Medical School. Thanks to Phil, MBT really thrived during our first four years at the UW (1992–1996).

I was beginning to realize that although a cross-disciplinary environment represented an essential foundation for systems biology, systems biology required many additional cultural changes quite at odds with the traditional practice of biology and its governance by traditional academic bureaucracies. In addition to a superb cross-disciplinary environment, one had to create a culture of teamwork for hard problems, high-throughput facilities for biological measurements, a strong computational infrastructure and the ability to create strategic partnerships with academia, research institutes and industry to fill in missing biological and technical skills. The cross-disciplinary culture required that the scientists learn one another's languages, that the nonbiologists learn relevant biology deeply, and that the scientists effectively work together in teams with continual feedback and interactions. One had to develop expensive high-throughput genomic and proteomic facilities (requiring space,

capital investment, and ongoing support) and to possess the ability to manage these rapidly changing technologies and keep them up to date. It was necessary to have extensive computational facilities. As noted above, one had to carry out team science, which led to potential difficulties with tenure for younger faculty. One needed to be able to readily make outside strategic partnerships to bring in needed engineering or scientific expertise for hard problems. Clearly, systems approaches were going to spin off an enormous amount of intellectual property, and managing that intellectual property in a manner that could successfully create additional support for the science was essential. It was also clear that extensive fundraising was necessary to enable this vision (but this was impossible in the context of a large state university where presidents and deans jealously guard the right to raise funds from private individuals). Each of these issues posed serious problems for a traditional academic bureaucracy.

Here I present two examples of the challenges faced in building systems biology at the UW. First, at Caltech I had helped raise the money to build two biology buildings; at the UW I was never once asked to do significant fundraising, apart from my initial interactions with Bill Gates, whose support created MBT. Yet, systems biology clearly was going to require significant resources. Second, the MBT cross-disciplinary environment needed to be significantly expanded. This proved to be challenging: For example, the new dean refused to let me hire a wonderful surface chemist from Penn State who could have contributed enormously to nanotechnology measurement approaches for systems biology, approaches we have subsequently collaborated on with Jim Heath at Caltech (discussed below). The dean claimed that surface chemistry was irrelevant to the medical school; this was, once again, a perfectly reasonable position for someone who did not understand how changing technologies would transform biology and medicine.

After trying to compromise on many of these issues, it became obvious that the administrative structure of the university could not accommodate most of our requirements (the workings of bureaucracies are honed by past experience and can rarely accommodate future change with ease). Hence, I decided in late 1999 to resign from the UW and start the independent Institute for Systems Biology (ISB), one of the first, if not the first, systems-biology centers in the world. After raising some funds, I persuaded Alan Aderem (an immunologist) and Ruedi Aebersold (a protein chemist and pioneer in proteomics) to join me as cofounders of ISB, which started in early 2000. These scientists, together with ISB faculty member John Aitchison, have played fundamental roles in the emergence of systems biology and/or technology development at the institute. But before discussing ISB, I here summarize the record of MBT.

5.4. Molecular Biotechnology's Remarkable Record of Accomplishment

MBT, arguably the first truly cross-disciplinary biology department in the United States, was strikingly successful. Ruedi Aebersold and John Yates advanced the field of proteomics by developing, respectively, an isotope-labeling technique known as isotope-coded affinity tags that permitted relative or absolute quantification of

proteins by mass spectrometry (81) and an important computational technique called Sequest that allowed proteins to be identified from a database search of tryptic peptides present in all proteins (as determined computationally from complete genome sequence) against the actual mass spectrometry measurements (82). Phil Green developed two computational tools that proved essential for the human genome project: (*a*) a tool that allowed short sequenced DNA fragments to be assembled into larger fragments, or contigs, and (*b*) software that assessed the quality of DNA sequence data (83). Ger van den Engh pioneered the development of a multiparameter, very high speed cell sorter. Maynard Olson and I each directed 2 of the 16 international genome centers sequencing the human genome—and in this regard, Lee Rowen did a wonderful job managing the Hood center's efforts to sequence significant portions of human chromosomes 14 and 15 (as noted above) together with Anup Madan and Shizen Qin. As mentioned above, Alan Blanchard developed the ink-jet array technology. Debbie Nickerson began her pioneering work on SNP analyses. Barbara Trask pushed the development of technologies for chromosomal in situ hybridization (gene localization) so essential to human genetics.

After I resigned, MBT subsequently merged with the Department of Genetics to create the new Department of Genome Sciences. Some MBT faculty stayed at the university, and others (generally those who were more more technically oriented) moved on.

6. THE INSTITUTE FOR SYSTEMS BIOLOGY YEARS: 2000–PRESENT

As I discuss in this section, ISB embodies much of my philosophy on doing science and really represents the summation of much of what I have learned in my career. Hence, I take the liberty of discussing in more detail its historical context, its rationale, its culture, and how systems biology drove the development of new technologies and computational tools in my lab.

6.1. General Comments about Systems Biology in the Twenty-First Century

Let me set the general context for how I think about systems biology and its central role in twenty-first-century biology. Biology will be a dominant science in the twenty-first century—just as chemistry was in the nineteenth century and physics was in the twentieth century—and for a fascinating reason. The dominant challenge for all the scientific and engineering disciplines in the twenty-first century will be complexity, and biology is now in a unique position to solve the deep problems arising from its complexity and to begin to apply this knowledge to the most challenging issues of humankind. Biology will use systems approaches (holistic, as opposed to atomistic) and powerful new measurement and visualization technologies, as well as the new computational and mathematical tools that are emerging in the aftermath of the human genome project and the emergence of systems biology. Biology will make use of the fact that our models of biological complexity can be tested by experimentation.

Biology will also use the emerging insight that it can be viewed as an informational science (an idea articulated in References 71 and 84–88).

Solving the complexities of biology will enable scientists to achieve two important objectives: (*a*) Biology will begin to solve some of humankind's most challenging problems, including health care for all, agriculture, nutrition, and bioenergy. (*b*) It will bring to the other scientific and engineering disciplines solutions to many of their most vexing problems, such as integrative computing strategies to computer science, striking new chemistries to chemistry, molecular-level machines for manipulating matter and measurements to engineering, new materials to material sciences, new ways of thinking about complexity to physics, and new ways of deducing relevant historical pasts to geology and archeology. Thus, twenty-first-century biology will enrich most of the other scientific and engineering disciplines. This biological treasure trove of knowledge exists because biology has had three billion years of evolutionary trial and error to create, test, and perfect these scientific strategies and engineering solutions.

6.2. Biology as an Informational Science and the Institute for Systems Biology

ISB, created in 2000, articulated the vision of transforming modern biology by pioneering the systems approaches to decipher biological complexity and by viewing biology as an informational science. ISB argued that through the convergence of comprehensive systems approaches to biology (combining both holistic and reductionist approaches), the development of new technologies, and the creation of powerful new computational/mathematical tools, the complexity of biology could be penetrated and understood. ISB sought to pioneer and integrate approaches and tools for each of these areas.

The view of biology as an informational science provides a powerful conceptual framework for dealing with complexity. First, there are two fundamental types of biological information, the digital information of the genome and the outside or environmental information that impinges on and modifies the digital information. Indeed, this digital core of knowable information distinguishes biology from all the other scientific disciplines: None of the others have this digital and hence readily decipherable core of information. Systems biology attempts to understand the integration of the digital and environmental information that mediates the three fundamental processes of life: evolution, development (e.g., in humans growing from one cell in the fertilized egg to 10^{14} cells of many different types in the adult), and physiology (e.g., the immune response to an infection). Hence, the purpose of systems biology is to integrate the information from the digital genome and the environment to understand how life unfolds.

Second, biological information is captured, processed, integrated, and transferred by biological networks—interacting sets of RNAs, proteins, the control regions of genes, and small molecules—to the simple and complex molecular machines that actually execute the functions of life. Thus, a central focus of ISB is an understanding of the dynamical operation of biological networks in the context of evolution,

development, physiology, or even disease, as is the understanding of the construction and function of molecular machines. Third, biological information is encoded by a multiscale information hierarchy: DNA, RNA, proteins, interactions, biological networks, cells, tissues and organs, individuals, and, finally, ecologies. Importantly, the environment impinges upon each level of the hierarchy and modulates the digital informational output from the genome. Hence, to understand how systems operate at a particular level—say understanding the 50 or so proteins that mediate the cell cycle—one should capture in a global manner each level of information that lies between the phenotypic measurements (features of the cell cycle) and the core digital genome. The information at each level should then be integrated in such a manner that the environmental modifications are identified so as to understand how they impact the functioning of the systems.

Note that each of these levels of information poses chemical and technical challenges for their global analysis; this allows us the possibility to bring in emerging technologies, such as microfluidics, nanotechnology, in vivo and in vitro molecular imaging, and new chemistries for creating protein capture agents. For example, we need to develop many new global chemistries to study proteins as they dynamically execute their functions; these measurements include their structure, expression patterns, chemical processing and modification, half-lives, interactions with other informational molecules and small molecules, locations in the cell, and dynamically changing three-dimensional structures. Obviously, many different aspects of chemistry will play a vital role in creating these new tools of systems biology. The capture, validation, storage, analysis, integration, visualization, and graphical or mathematical modeling of data sets—dynamically captured and global in nature—pose a host of computational and mathematical challenges. This intellectual context, and the formation of the modern version of what we call systems biology, is at the heart of the creation of ISB and its work over the past seven years. The implementation of the scientific program that has emerged from this context is at the heart of the future of ISB, and requires an organizational and cross-disciplinary cultural context to assist in its development.

We have had the opportunity to pioneer systems approaches, genomic (75), proteomic (94–96), and single-cell technologies and a wide variety of computational tools essential for systems approaches (77, 84, 89–93, 98–100). The Hood lab has been involved in many aspects of systems biology, such as systems approaches to biology (76, 77, 84, 87, 103), disease (104–109), proteomics (110, 111), technology development (77–109), and the development of relevant computational and mathematical tools (84, 99, 112–124). Trey Ideker, one of my former graduate students and now a professor at the University of California at San Diego, played a special role in catalyzing the experimental beginnings of a systems approach to galactose utilization in yeast and in pioneering the development of the systems-biology standard for graphical network visualization, the software program Cytoscape (86; in my view one of the pioneering papers of systems biology).

ISB has established effective facilities for high-throughput data generation (genomic, proteomic, and single cell). In general, the frontier challenges of the biology that the ISB faculty are studying have driven all that ISB does.

6.3. The Culture and Philosophy of the Institute for Systems Biology

Some powerful new ideas that have begun to effectively integrate biology, medicine, technology, and computation/mathematics are key to the development of ISB as an institution. The first three of these ideas are fundamentally scientific; the rest are strategic and guide the implementation of the institution's development. Some of these points partially repeat earlier discussion, but I include them here for the sake of completeness.

1. The frontier problems of biology should dictate what technologies should be developed (making it possible to view new dimensions of biological data space). Likewise, the creation of new data sets from these technologies should drive the development of new computational and mathematical techniques for analyzing them. New technologies and computational tools in turn enable the understanding of new levels of biological complexity. Thus, the dynamics of emerging new technologies frame the rate at which new biological insights are generated.
2. This integration of biology, technology, and computation necessitates the creation of a cross-disciplinary environment that brings biologists, chemists, computer scientists, engineers, mathematicians, physicians, and physicists together to learn one another's languages and work together in teams and allows the nonbiologists to learn relevant biology in a deep manner. ISB is still struggling to achieve these objectives.
3. Biology spans a spectrum of complexity, ranging from simpler model organisms such as single-celled bacteria or yeast to more complex model organisms such as mice and, ultimately, humans. Most importantly, biological experimentation is easier in simple species. Hence, ISB develops new tools and approaches using simpler model organisms and then must learn how to apply these tools to higher model organisms—and ultimately to human complexity.
4. ISB is deeply committed to an open-source philosophy, that is, making our data and tools readily available to the scientific community and taking advantage of the collective input of this community to improve those tools.
5. ISB is committed to remaining small, focused, and highly interactive. To compensate for the limitations this philosophy imposes upon faculty size, ISB requires an immediately available reservoir of expertise. This critical mass of knowledge in relevant biology, chemistry, computer science, engineering, mathematics, and medicine resides with our 25 senior scientists as well as our 13 faculty members.
6. One of ISB's approaches to attacking big scientific problems while remaining relatively small is to create strategic partners that bring us the scientific, technological, computational, and medical expertise that we lack. We focus on partnering with the best. ISB searches for these partners among academics, industry, and research institutions, as well as in relevant institutions in foreign countries.
7. ISB is determined to transfer its relevant knowledge to society, be it through K–12 science education, spin-off companies employing ISB's novel biological or technology insights, or educating society in science and technology.

8. ISB believes in providing the scientific leadership necessary for catalyzing paradigm changes in how biology and medicine are carried out and in how biology and medicine are organized—and we have a long history of doing so (see Section 6.11 below).

The following examples provide a glimpse into the ways the Hood lab at ISB plays a leading role in catalyzing the emergence of new organizational structures and creating powerful strategic partnerships.

6.4. Prostate Cancer

At the UW, my work turned to more medically focused problems, and I continued and extended those efforts at ISB. I took a systems approach to cancer biology, with a particular focus on prostate cancer—both from the viewpoint of genome-wide genetic mapping (125–133), which we carried out with our wonderful collaborators Elaine Ostrander and Janet Stanford at the Fred Hutchinson Cancer Research Center, and from the viewpoint of using DNA array technology to understand the dynamics of gene expression in the cancer disease process (134–144). This latter effort included productive collaborations with a prostate cancer clinician, Paul Lange at the UW; an MD postdoctoral fellow in my lab, Pete Nelson (now on the faculty at the Fred Hutchinson Cancer Research Center); and Biaoyang Lin, a senior scientist and long-time colleague at ISB. This project really drove the application of DNA array technologies (see, e.g., Reference 109): It pioneered applications using the then-new digital transcript counting technology known as multiple parallel signature sequencing (144), and we (in collaboration with the biotech company Helicos) have begun to explore the use of the powerful new single-strand DNA–sequencing technology to study interesting tissue transcriptomes.

In the context of the genetic mapping prostate project, I learned an important lesson about thinking outside the box. I received support from Michael Milken's Prostate Cancer Foundation and got to know Michael reasonably well. I remember in the mid-1990s talking with him about the challenge of obtaining families with a history of prostate cancer—collecting sufficient numbers of such families to provide adequate statistical power for the genetic analyses often took clinicians 10–20 years. Michael's solution was typically simple, rapid, and effective. He proposed that he and I, together with General Norman Schwarzkopf (who had prostate cancer, as had Michael), go on Larry King Live and make an hour-long pitch for these prostate-cancer families. I initially thought this idea was ridiculous. However, we did the program with Larry King, which was one of his most popular that year (1995). In six weeks we had recruited about 250 prostate cancer families from around the world (most from the United States), and together with wonderful collaborators—Elaine Ostrander and Janet Stanford at the Fred Hutchinson Cancer Research Center—we collected appropriate DNA samples from the families, then carried out a long-term series of genetic mapping studies.

The King show was a brilliant and imaginative solution to the challenging problem of disease-family collection. Frustration with the limited success of these

family studies led us at ISB to begin thinking about a new genetic approach to understanding disease; we term this approach systems genetics (see Section 6.5). More recently, we have begun analyzing the biology and therapeutic responses of human glioblastomas and ovarian cancer using these same powerful genomic and proteomic technologies.

These cancer biology studies are driving tool development in several respects: (*a*) software algorithms to statistically assess and integrate different data sets both of the same data type and of different data types (112); (*b*) computational techniques for visualizing the dynamics of disease-perturbed networks, and ultimately their graphical or mathematical modeling; (*c*) new computational approaches to reducing the enormous data dimensionality of DNA array studies to simple hypotheses about health and disease (see Reference 115 for a fascinating successful example); and (*d*) high-throughput (digital) DNA-sequencing methods (e.g., single-stranded DNA sequencing) to quantitatively delineate dynamically changing transcriptomes. Ultimately, rapid DNA sequencing will completely replace DNA arrays in this task and will eventually determine the sequences of individual genomes (see below).

6.5. Systems Genetics

The genome-wide genetic mapping results on prostate cancer were frustrating in that, as is generally the case in the genetic studies of complex diseases, signal-to-noise problems posed significant challenges. I assume that for most complex genetic diseases there may be 30 or more potential variant genes involved in a combinatorial manner in smaller subsets, for example, where the appropriate six or so genetic variants can work together in various combinations to generate the distinct disease phenotypes. With most genome-wide association studies now being carried out on hundreds of thousands to millions of SNPs and thousands of patients, in conjunction with the recently completed human haplotype map, the signal-to-noise ratio is so poor that experimenters are fortunate if they can identify even one dominant gene out of the many involved in the complex disease process.

Of course, this problem can be somewhat mitigated by combining the association studies with other types of genomic data (transcriptome quantitation, indels, amplifications, or deletions of genomic regions, etc.). However, we feel that an entirely new approach is required for high-resolution disease-gene finding in complex genetic diseases. Indeed, one of the efforts that is just now getting under way at ISB is the creation of a new field we call systems genetics, wherein we attempt to connect genotype and phenotype together through an understanding of biological network behaviors of relevant variant genes. We are beginning these systems-genetics efforts using yeast as a simple model system (pioneered by ISB faculty members Aimee Dudley, Tim Galitski, John Aitchison, and David Galas). These efforts are driving us to develop (or to collaborate with those developing) two emerging technologies: (*a*) very-high-throughput DNA sequencing, which should allow the determination of an individual's genome sequence quickly and for less than one thousand dollars within the next ten years with the use of one or more of the emerging next-generation

sequencing strategies, and (*b*) the detailed dynamic analysis of thousands of single cells obtained from an individual and perturbed by hundreds of appropriate environmental agents to interrogate differing subsets of the function networks of individual cells to reveal the activity of the underlying biological cellular networks in the context of an individual's known genome sequence. This analysis uses the microfluidic approaches that Steve Quake has pioneered with soft polydimethylsiloxane (PDMF) materials, allowing one to create miniaturized valves, pumps, and mixing chambers, which are needed to analyze fluids and cells. Perhaps we can do systems genetics by looking at what single cells can tell us if appropriately perturbed. We are working together with strategic partners in each of these important areas.

6.6. Neural Degenerative Diseases

I have also become interested in applying a systems approach to the study of neural degenerative diseases. We have been studying prion disease in mice with George Carlson (a wonderfully interactive Montana colleague) and Stanley Prusiner for the past 25 years, and more recently with Inyoul Lee, a senior scientist in my lab.

Recently we looked at mice infected intercranially with infectious prion proteins. We studied the dynamic brain transcriptomes (populations of mRNAs in the brain) in five different mouse inbred strains, two congenic strains, and one knockout strain at ten different time points across the progression of their prion disease. We studied the large number of different strains because each serves as a fascinating biological filter to deal with the signal-to-noise challenges of transcriptome data; this strategy allowed us to identify and assess the core set of genes encoding the prion disease process. We compared and integrated the dynamically changing disease and normal control brain transcriptomes to identify the key genes that have changed activity as a consequence of the disease. We then integrated the transcriptome data and mapped them onto known biological networks together with phenotypic data (the histopathology and the clinical signs). These studies led to several interesting conclusions. For instance, the dynamically changing brain networks can explain much of the pathophysiology of prion disease, but these networks also provide a new systems approach to thinking about blood diagnostics. Systems medicine is predicated on the simple idea that disease arises from one or more disease-perturbed networks in the affected organ (perturbed genetically and/or environmentally) (108). This alters the patterns of dynamically expressed information from these networks, and these altered dynamically changing patterns of transcription encode the corresponding dynamically changing pathophysiology of the disease. Systems medicine is driving the development of new measurement technologies [DNA sequencing to measure genomes and transcriptomes; protein-measurement technologies (see below); single-cell analyses of DNA, RNA, proteins and protein-protein, protein-DNA and protein–small molecule interactions; and in vivo and in vitro molecular-imaging technologies] to determine how networks are changing in individual cells or individual organisms. Once again, ISB is itself developing several of these technologies and is collaborating with strategically chosen partners on others.

6.7. Organ-Specific Blood Protein Fingerprints: A Systems Approach to Disease Diagnostics

I describe the systems approach to disease diagnostics in detail because I believe it is going to be one of the most transformational approaches in the new medicine and because it beautifully embodies the essence of systems medicine. The dynamically changing prion networks revealed that levels of some transcripts were elevated (or decreased) many weeks before the detection of clinical signs. If some of these transcripts encode proteins that are secreted into the blood, then their elevated levels in blood might be an early preclinical sign of incipient disease. With proteomics discovery approaches, it is relatively easy to identify quantitative changes in levels of many proteins that distinguish, for example, individuals with ovarian cancer from their healthy counterparts. However, if these same markers are examined in the blood of individuals with, say, ten other diseases, they behave in unpredictable ways because these markers are generally synthesized in multiple organs and hence are responsive to different environmental stimuli.

Our idea has been to use transcriptome analyses to identify the organ-specific transcripts in, for example, the brain, by comparing the brain transcriptome against the transcriptomes of 40 other organs and determining which transcripts are primarily expressed in the brain. If this is done with every organ and tissue, we anticipate that most will contain 100–200 organ-specific transcripts. If the protein products of some of these transcripts are secreted into the blood, they constitute an organ-specific blood protein fingerprint wherein the levels of the individual proteins reflect the operations of the corresponding organ networks that encode them. Accordingly, a normal individual will have one set of levels of brain-specific proteins in his or her blood fingerprint, whereas the levels of some of these proteins will change in ways that are specific to each different brain disease (i.e., brain cancer or brain infection) because each is encoded by different combinations of disease-perturbed networks. Because we would like to deduce the nature of the disease-perturbed networks from the dynamically changing concentrations of proteins in the organ-specific fingerprint, we need to measure at least 50 proteins per fingerprint. The technical challenge, then, is twofold: (*a*) One must measure 50 proteins for the organ fingerprints for each of the 50 or so different human organs (2500 measursements), and (*b*) the assay should be carried out from a single droplet of blood. Thus, this measurement technology must be miniaturized and highly parallelized: Technology using microfluidic and/or nanotechnology approaches appears to be required.

The analysis of biomarkers in the blood represents an enormous technical challenge to proteomics. Blood is a mixture of millions of proteins whose concentrations span a dynamic range of perhaps 10^{12}, with 21 proteins constituting approximately 99% of the blood protein mass. [Keep in mind we are talking only about the identification and quantification of individual biomarkers (proteins). There remain enormous chemical challenges with regard to developing techniques that can characterize the additional types of protein diversity that arise posttranscriptionally, delineate different forms from alternative RNA splicing, characterize the processing of proteins by enzymes, detect the 400 or more different potential chemical modifications of proteins,

and measure the different half-lives of proteins. In addition, proteins change their structures dynamically during the execution of their functions and migrate to different regions of the cell to carry out specific functions; global technologies to measure these features also need to be developed.] We need to quantify proteins secreted from large organs (e.g., the liver) and small organs (e.g., the prostate); hence, we need measurement techniques for proteins that span a very large dynamic range in the blood.

The challenge to the identification of blood biomarkers is twofold. First, one must discover appropriate biomarkers (in our case, the organ-specific proteins), which requires the analysis of tens to hundreds of samples. This can be done by a variety of techniques including mass spectrometry and antibody-based assays (enzyme-linked immunosorbent assay, Western blot, surface plasmon resonance, etc.). My lab is now collaborating with the biotech company Plexera to develop a surface plasmon resonance instrument that can measure 800 different antibody interactions in just five minutes and can repeat the cycle without signal degradation every ten minutes for up to 40 cycles—another example of a key high-throughput measurement technology, in this case for proteomics. (Ruedi Aebersold has been a cutting-edge pioneer in developing a wide variety of proteomics techniques and strategies for analyzing biomarkers; see References 92–95.) Also, there remain striking chemical challenges for the development of these discovery approaches, which need to be more sensitive, more global, and more specific. Second, one must eventually be capable of large-scale typing—perhaps the quantification of 2500 blood proteins from one droplet of blood in hundreds of millions of patients per year. Typing mandates the use of microfluidic and nanotechnology measurement strategies.

Four years ago, I started a collaboration with Jim Heath, a young chemist at Caltech, using microfluidic and nanotechnology approaches to quantify organ-specific (and other) blood proteins. Jim's lab has recently developed a new type of protein chip, known as a DNA-encoded antibody library (145). This protein chip appears to have a dynamic range of 10^8 and a sensitivity in the low femtomole range, and it can potentially be manufactured inexpensively in large quantities. We currently have protein chips with approximately 20 protein-capture features—a scale that can eventually be expanded to thousands of features. This feature scale is limited only by the availability of protein-capture agents with high specificity and affinity (currently antibodies, but we are exploring alternative chemical possibilities for synthesizing highly specific reagents). This collaboration has progressed rapidly over the past four years and illustrates the power of carefully chosen strategic partnerships wherein all parties bring together complementary scientific skills. The collaboration with Jim has been one of the best I have ever had.

6.8. Predictive, Personalized, Preventive, and Participatory (P4) Medicine

The convergence of systems approaches to disease, new measurement and visualization technologies, and new computational and mathematical tools suggests that our current largely reactive mode of medicine (i.e., wait until one is sick before responding) will over the next 10–20 years be transformed to predictive, personalized,

preventive, and participatory (P4) medicine (104–106, 108). Two components of predictive medicine will emerge over the next ten years: (*a*) Individual genome sequences will be available, and (*b*) increasingly, we will be able to determine the likelihood of one's future health (e.g., 50% probability of ovarian cancer by age 50). Hand-held devices to prick the finger and quantify 2500 organ-specific proteins from all human organs will send this information via wireless communication to a server, which in turn will analyze the information and email a report to the patient and their physician. This rapid communication, done perhaps twice a year, will thus permit an instantaneous assessment of current health status. These measurements will themselves personalize medicine. And we must remember that each of us differs on average by approximately six million nucleotides from our neighbors; hence, we are susceptible to differing combinations of diseases and, once again, must be treated as individuals.

From the assessment of genomes and environmental exposures will emerge initially a predictive and personalized medicine. Later, physicians will learn how to identify drugs to re-engineer disease-perturbed networks, causing them to behave in a more normal fashion, or at least abrogating the most deleterious of their effects. In the future, we will be able to design drugs to prevent networks from becoming disease perturbed. For example, if there is an 80% change of prostate cancer at age 50, taking these preventive drugs beginning at age 35 may reduce disease probability to 2%. Finally, if we can educate patients and their physicians as to the nature of P4 medicine, then patients will be in a position to take more responsibility for charting and participating in their own future health choices. The realization of P4 medicine is a major strategic goal of ISB.

The vision of P4 medicine has emerged from ISB, but it has also emerged from a strategic partnership (described above) that brings together three critical skills: (*a*) systems biology and medicine (Hood and Galas, ISB); (*b*) microfluidics and nanotechnology (Heath, Caltech); and (*c*) molecular imaging (Mike Phelps, inventor of positron emission tomography scanning, University of California at Los Angeles). This partnership, termed the NanoSystems Biology Alliance, has facilitated the creation of a series of NIH centers (e.g., the Systems Biology Center at ISB and the NanoSystems Biology Cancer Center at Caltech). Our P4 vision has been delineated in a series of papers (104, 106–108) and has been a stimulating and broadening opportunity for all involved. ISB recently has elected to make P4 medicine one of its central strategic projects, and all of its faculty are now in discussion about the convergence and focus of our collective talents on this challenging problem.

6.9. P4 Medicine and Its Implications for Society

P4 medicine has several fascinating implications. It will, over the next ten years, transform the business plans of virtually every sector of the health care industry—pharmaceuticals, biotech, medical instrumentation, diagnostics, health care information technology, payers, providers, medical centers, medical schools, and so forth. For example, pharmaceutical companies are generally acknowledged as failing in their quest to produce effective and reasonably priced drugs. Systems medicine will bring presymptomatic diagnostics and the ability to stratify disease so that effective

therapies can be successfully matched against specific diseases. It will provide powerful new approaches to assessing drug toxicity early in clinical trials. It will also provide new approaches to assessing drug doses for individual patients and evaluating drug toxicities at a very early stage. A fascinating question is whether the pharmaceutical companies will be able to effectively employ these strategies of systems medicine. Another challenging issue for P4 medicine concerns medical schools, which are currently teaching physicians that will be practicing P4 medicine in 10–20 years. However, these students are not learning the background and concepts that they will need for P4 medicine. Will medical schools be able to transform their teaching, research, and eventually their patient care to encompass the P4 concepts? Similar issues apply to every sector of the health care industry.

P4 medicine will lead to the digitalization of medicine, that is, the ability to extract disease- or health-relevant information from single molecules, single cells, or single individuals. The digitization of medicine will have a far greater impact on society than will the digitization of communications or information technology. The reason for this is because at some time in the future (depending upon the rate at which technologies emerge and the extent of federal and private resources that can be focused on P4 medicine), there will be a sharp turnaround in the ever-escalating costs of health care to the point that we will be able to export P4 medicine to the developing world. Indeed, P4 medicine will, in the near future, form the foundation of global medicine.

P4 medicine poses significant technical and social challenges that are amenable to powerful cross-disciplinary scientific attack and societal education and debate. The societal challenges must be dealt with at the same time the technical challenges are being solved if P4 medicine is to successfully emerge in the next 10–20 years.

ISB is in the process of generating strategic partnerships to attack the technical and societal challenges to P4 medicine. My collaborations with David Galas and Diane Isonoka at ISB have led to a series of fascinating possibilities for strategic partnerships—with individuals, academic centers, companies, and even with countries—that will not only bring critical scientific, engineering, and clinical expertise, but will also provide striking new funding opportunities, some of which are global in nature. Indeed, I envision a globalization of science emerging over the next ten years or so—just as Tom Friedman described the globalization of the economy. And in a similar manner, there are striking opportunities for those who will be at the leading edge of scientific globalization.

6.10. Using Intellectual Property to Support Science: The Accelerator

About three years ago I decided to create a for-profit company called the Accelerator, whose mission was to create successful new biotech companies with the help of ISB. Those companies, in turn, could generate the resources for a long-term ISB endowment. Venture capital companies at that time were focusing primarily on companies generating late-phase drugs, and I wanted to facilitate the emergence of new companies with large-scale discovery platforms or new strategies for extracting relevant

biological information for medicine. I went to about 20 venture capital companies with a proposal for a group of five or so venture capitalists to join with ISB to create the Accelerator. Each venture group would contribute $3–5 million. A board would be established with representative membership from each venture company, the CEO of the Accelerator (Carl Weissman, who has done a superb job), and myself. This board would select suitable companies to support for two to three years to prove the principles of their scientific approaches. Then the new company would graduate by raising series B (second-round) money and would go out on their own—or, if the company could not raise money after a few years, it would be terminated. The Accelerator would handle all the management for each company.

We also recruited Alexandria, a real estate company specializing in laboratory construction. They built beautiful facilities for us at a reasonable cost with the expectation of providing space for some of the successful companies as they graduated from the Accelerator. ISB played a special role in creating the Accelerator, as well as in providing faculty support for due diligence and the scientific advisory boards, ISB's excess capacity for high-throughput genomic and proteomic measurements, access to our outstanding computational facilities, and my expertise of almost 30 years in starting companies. For this, ISB receives an equity position in each company that will be used to build our endowment in the next 10–15 years; some of these companies will hopefully mature to become as successful as some of the past companies we have founded. To date we have screened more than 400 business plans—those coming from the venture companies as well as those identified by Carl Weissman and myself—and we have selected six. The results have been outstanding. Three of the companies have successfully raised series B money—from $30 million to $55 million. We believe that two of the remaining three companies will also do very well—a remarkable record when compared with most venture efforts.

Recently we raised $22.5 million for the second Accelerator round, with the expectation that we will bring five or six new companies to the Accelerator over the next few years. Indeed, in the last year and a half, the venture money raised in the state of Washington for biotech has been approximately $160 million—and the Accelerator has raised more than 70% of this money. This testifies to what a small, focused, and knowledgeable effort can do. This will probably be a far more successful approach to converting scientific knowledge into support for science than the vast majority of the efforts directed at licensing intellectual property to pre-existing companies. Indeed, in my experience, only the licensing of the automated DNA sequencer to ABI generated significant income (on the order of $100 million for Caltech), and it was the most successful licensing in Caltech's history. But significant financial licensing success is rare indeed.

6.11. Paradigm Changes, or New Ideas Often Need New Organizational Structures for Their Realization

Whether instrument integration and high data throughput, the human genome project, cross-disciplinary biology, systems biology, and P4 medicine are really

paradigm changes or just interesting new ideas is partly in the eye of the beholder. What I can say for certain is that each of these concepts was initially greeted with considerable skepticism from the scientific or medical communities. I believe each of these innovations has impacted biology in major ways, and the first four required new organizations to begin to realize their potential. We do not yet know what types of new organizations P4 medicine will require for its realization. I stress now, as I did at the beginning of this article, that I am but one player of many who participated in each of these five paradigm changes. I summarize my experiences as follows:

1. To realize the full development, commercialization, and widespread application of the first four chemical technologies that we developed, we needed to set up a new organization: ABI. We came to this realization with striking clarity when Bob Sinsheimer, then Chairman of Biology at Caltech, stopped by my office in mid-1973 to suggest in the strongest possible terms that I give up technology development and focus only on biology. This did not have to do with tenure, as I was awarded tenure just a few months later (quite early for Caltech). Rather, it appeared to have to do with taste. I refused, of course, and 20 years later finally asked Bob why he had paid me that visit. He said the senior biologists at Caltech felt it was inappropriate to mix engineering (and probably commercialization) with biology. I later created ABI to take on much of this engineering and commercialization.
2. When I (along with several others) first went into the biological community in 1985 to push the human genome project, most biologists and NIH were firmly opposed. Rationality eventually held sway, but only after a new NIH institute, the National Human Genome Research Institute, was created to manage the genome project.
3. When I decided to build a cross-disciplinary biology department, I had to move from Caltech to the UW to create MBT. As illustrated above, this new department—the first of its kind—flourished.
4. In building an environment for systems biology it once again became clear that a new and independent organization, ISB, was necessary to realize systems biology, and it has done so with considerable success.
5. Finally, it will be interesting to see how P4 medicine emerges. Will the established companies in the health care sector be capable of the transformational change required to take advantage of the new opportunities emerging from systems medicine and a systems approach to disease? Or will many new organizations emerge focused precisely and totally on the objectives of P4 medicine with appropriate technologies? The jury is out, but I expect that many new health care companies will form over the next 10–20 years in response to the disruptive technologies and strategies emerging from P4 medicine. There is also the question of how the academic medical environment will embrace these new opportunities; doing so in an incremental fashion, as is usually the case, generally leads to modest evolutionary change and not the desired revolutionary changes. The future will be very exciting.

7. A FEW CLOSING COMMENTS FOR STUDENTS

I leave students (and even some of my colleagues) with several pieces of advice. First, I stress the importance of a good cross-disciplinary education. Ideally, I suggest a double major with the two fields being orthogonal—say, biology with computer science or applied physics. Some argue that there is insufficient time to learn two fields deeply at the undergraduate level. I argue that this is not true. If we realize that many undergraduate courses now taught are filled with details that are immediately forgotten after the course is finished, we must then learn to teach in an efficiently conceptual manner. As I noted above, as an undergraduate at Caltech I had Feynman for physics and Pauling for chemistry, and both provided striking examples of the power of conceptual teaching. Second, I argue that students should grow accustomed to working together in teams: In the future, there will be many hard problems (like P4 medicine) that will require the focused integration of many different types of expertise. Third, I suggest that students acquire an excellent background in mathematics and statistics and develop the ability to use various computational tools. Fourth, I argue that a scholar, academic, scientist, or engineer should have four major professional objectives: (*a*) scholarship, (*b*) education (teaching), (*c*) transferring knowledge to society, and (*d*) playing a leadership role in the local community to help it become the place in which one would like one's children and grandchildren to live. Fifth, with regard to the scientific careers of many scientists—they can be described as bell-shaped curves of success—they rise gradually to a career maximum and then slowly fall back toward the base line. To circumvent this fate, I propose a simple solution: a major change in career focus every 10 or so years. By learning a new field and overcoming the attendant insecurities that come from learning new areas, one can reset the career clock. Moreover, with a different point of view and prior experience, one can make fundamental new contributions to the new field by thinking outside the box. Then the new career curve can be a joined series of the upsides of the bell-shaped curve, each reinvigorated by the ten-year changes. Finally, science is all about being surrounded by wonderful colleagues and having fun with them, so I recommend choosing one's science, environment, and colleagues carefully. I end this discussion with what I stressed at the beginning—I am so fortunate to have been surrounded by outstanding colleagues who loved science and engineering (**Table 1**). Science for each of us is a journey with no fixed end goal. Rather, our goals are continually being redefined.

8. PERSONAL THOUGHTS ABOUT MY CAREER EXPERIENCES

In retrospect I see the 40 years of my career as embodying and leading toward the same principles that we brought to ISB. But at the very foundation of all the science and technology I have enjoyed are the wonderful colleagues I have associated with throughout my career—they are the true pioneers of virtually all that has been accomplished (**Table 1**). Biology is the central focus and driver of the technologies and strategies to be developed. Indeed, I see my career as having pointed toward the creation of the cross-disciplinary environments that enabled this virtuous cycle

of biology driving technology, and technology in turn generating data that drive the development of computational and mathematical tools. Thus biology drives both technology and computation.

Also, I see myself moving throughout much of my career toward a systems approach, first to biology and then, more recently, to medicine. Each of the technologies that we developed focused upon solving chemical problems—for the essence of life is the chemistry of how living organisms deal with biological information. Another major aspect of my career has been to participate in the creation of new visions for how biology should be carried out—and I was often able to create new organizational structures that could allow these visions to manifest themselves. I wonder what new organizational structures for practicing both biology and P4 medicine will emerge over the next ten years. What a fascinating time to be in science and technology!

DISCLOSURE STATEMENT

The author is not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I would like to thank Lee Rowen for her critical review of this manuscript and always-insightful feedback and Shirley Meinecke for assisting in the manuscript's production.

LITERATURE CITED

1. Dyson FJ. 1998. *Imagined Worlds*. Cambridge, MA: Harvard Univ. Press
2. Hood LE, Gray WR, Dreyer WJ. 1966. On the mechanism of antibody synthesis: a species comparison of l-chains. *Proc. Natl. Acad. Sci. USA* 55:826–32
3. Hood L, Gray WR, Dreyer WJ. 1966. On the evolution of antibody light chains. *J. Mol. Biol.* 22:179–82
4. Gray WR, Dreyer WJ, Hood L. 1967. Mechanism of antibody synthesis: size differences between mouse kappa chains. *Science* 155:465–67
5. Hood L, Gray WR, Sanders BG, Dreyer WJ. 1967. Light chain evolution. *Cold Spring Harbor Symp. Quant. Biol.* 32:133–46
6. Silver J, Hood L. 1975. Automated microsequence analysis by use of radioactive phenylisothiocyanate. *Anal. Biochem.* 67:392–96
7. Hunkapiller MW, Hood L. 1980. New protein sequenator with increased sensitivity. *Science* 207:523–25
8. Hewick RM, Hunkapiller MW, Hood LE, Dreyer WJ. 1981. A gas-liquid-solid-phase peptide and protein sequenator. *J. Biol. Chem.* 256:7990–97
9. Doolittle RF, Hunkapiller MW, Hood LE, Devare SG, Robbins KC, et al. 1983. Simian sarcoma virus one gene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor. *Science* 221:275–77
10. Prusiner SB, Groth DF, Bolton DC, Kent S, Hood LE. 1984. Purification and structural studies of a major scrapie prion protein. *Cell* 38:127–34

11. Desch B, Westaway D, Walchli M, McKinley MP, Kent SBH, et al. 1985. A cellular gene encodes scrapie PrP 27–30 protein. *Cell* 40:735–46
12. Raftery MA, Hunkapiller MW, Strader CD, Hood L. 1980. Acetylcholine receptor: complex of homologous subunits. *Science* 208:1454–57
13. Hunkapiller MW, Hood LE. 1980. Human fibroblast interferon: amino acid analysis and amino terminal amino acid sequence. *Science* 207:525–26
14. Zoon KC, Smith ME, Bridgen PJ, Anfinsen CB, Hunkapiller MW, Hood LE. 1980. Amino terminal sequence of the major component of human lymphoblastoid interferon. *Science* 207:527–28
15. Goldstein AS, Tachibana L, Lowney I, Hunkapiller MW, Hood LE. 1979. Dynorphin (1–13), an extraordinarily potent opioid peptide. *Proc. Natl. Acad. Sci. USA* 76:6666–70
16. Schally AV, Huang WY, Chang RCC, Arimura A, Redding TW, et al. 1980. Isolation and structure of pro-somatostatin: a putative somatostatin precursor from pig hypothalamus. *Proc. Natl. Acad. Sci. USA* 77:4489–93
17. Heller EL, Kaczmarek K, Hunkapiller MW, Hood LE, Strumwasser F. 1980. Purification and primary structure of two neuroactive peptides that cause bag cell after discharge and egg-laying in *Aplysia*. *Proc. Natl. Acad. Sci. USA* 77:2328–32
18. Goldstein A, Fischli W, Lowney LI, Hunkapiller M, Hood L. 1981. Porcine pituitary dynorphin: complete amino acid sequence of the biologically active heptadecapeptide. *Proc. Natl. Acad. Sci. USA* 78:7219–23
19. Horvath SJ, Firca JR, Hunkapiller T, Hunkapiller MW, Hood L. 1987. An automated DNA synthesizer employing deoxynucleoside 3′ phosphoramidites. *Methods Enzymol.* 154:314–26
20. Strauss EC, Kobori JA, Siu G, Hood LE. 1986. Specific primer-directed DNA sequencing. *Anal. Biochem.* 154:353–60
21. Kent SB, Hood LE, Beilan H, Marriot M, Meister S, Geiser T. 1984. A novel approach to automated peptide synthesis based on new insights into solid phase chemistry. In *Proceedings of the Japanese Peptide Symposium*, ed. N. Isymiya, pp. 217–22. Osaka: Protein Res. Found.
22. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, et al. 1989. Structure of a complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. *Science* 246:1149–52
23. Clark-Lewis I, Hood LE, Kent SB. 1988. Role of disulfide bridges in determining the biological activity of interleukin 3. *Proc. Natl. Acad. Sci. USA* 85:7897–901
24. Clark-Lewis I, Lopez A, Luen B, Vadas M, Schrader JW, et al. 1989. Structure-function studies of human granulocyte-macrophage colony-stimulating factor: identification of amino acids required for activity, and an 84-residue active fragment. *J. Immunol.* 141:881–89
25. Pärraga G, Horvath SJ, Eisen A, Taylor WE, Hood L, et al. 1988. Zinc-dependent structure of a single-finger domain of yeast ADM. *Science* 241:1489–92
26. Maxam AM, Gilbert W. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* 74:560–64

27. Sanger F, Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 94:441–48
28. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, et al. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* 321:674–79
29. Kaiser RJ, MacKellar SL, Vinayak RS, Sanders JZ, Saavedra RA, Hood LE. 1989. Specific-primer-directed DNA sequencing using automated fluorescence detection. *Nucleic Acids Res.* 17:6087–102
30. Int. Hum. Genome Seq. Consort. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
31. Hood L, Rowen L. 1997. The impact of genomics on 21st century medicine. *Contemp. Urol.* 9:86–98
32. Hood L. 2002. A personal view of molecular technology and how it has changed biology. *J. Proteome Res.* 1:399–409
33. Hunkapiller M, Kent S, Caruthers M, Dreyer W, Firca J, et al. 1984. A microchemical facility for the analysis and synthesis of genes and proteins. *Nature* 310:105–11
34. Olson M, Hood L, Cantor C, Botstein D. 1989. A common language for physical mapping of the human genome. *Science* 245:1434–35
35. Wang K, Koop BF, Hood L. 1994. A simple method using T4 DNA polymerase to clone polymerase chain reaction products. *Benchmarks* 17(2):236–38
36. Wilson R, Chen C, Hood L. 1990. Optimization of asymmetric polymerase chain reaction for rapid fluorescent DNA sequencing. *BioTechniques* 8:184–89
37. Lai E, Wang K, Avdalovic N, Hood L. 1991. Rapid restriction map constructions using a modified pWE15 cosmid vector and a robotic workstation. *BioTechniques* 11:212–17
38. Huang GM, Wang K, Kuo C, Paeper B, Hood L. 1994. A high-throughput plasmid DNA preparation method. *Anal. Biochem.* 223:35–38
39. Wang K, Gan L, Boysen C, Hood L. 1995. A microtiter plate–based high-throughput DNA purification method. *Anal. Biochem.* 226:85–90
40. Suzuki M, Baskin D, Hood L, Loeb LA. 1996. Random mutagenesis of *Thermus aquaticus* DNA polymerase I: concordance of immutable sites in vivo with the crystal structure. *Proc. Natl. Acad. Sci. USA* 93:9670–75
41. Wang K, Boysen C, Shizuya H, Simon MI, Hood L. 1997. Complete nucleotide sequence of two generations of a bacterial artificial chromosome cloning vector. *Benchmarks* 23:992–93
42. Guo Z, Gatterman MS, Hood L, Hansen JA, Petersdorf EW. 2002. Oligonucleotide arrays for high-throughput SNPs detection in the MHC class I genes: HLA-B as a model system. *Genome Res.* 12:447–57
43. Hunkapiller MW, Hood LE. 1978. Direct microsequence analysis of polypeptides using an improved sequenator, a nonprotein carrier (polybrene), and high-pressure liquid chromatography. *Biochemistry* 17:2124–33
44. Hunkapiller MW, Lujan F, Ostrander F, Hood L. 1983. Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. *Methods Enzymol.* 91:227–36

45. Aebersold RH, Teplow DB, Hood LE, Kent SBH. 1986. Electroblotting onto activated glass: high efficiency preparation of proteins from analytical sodium dodecyl sulfate–polyacrylamide gels for direct sequence analysis. *J. Biol. Chem.* 261:4229–38
46. Aebersold R, Teplow DB, Hood LE, Kent SBH. 1986. Electroblotting from immobiline isoelectric focusing gels for direct protein sequence determination. *Peptides Biol. Fluids* 34:715–18
47. Kent S, Hood L, Aebersold R, Teplow D, Smith L, et al. 1987. Approaches to sub-picomole protein sequencing. *BioTechniques* 5:314–21
48. Aebersold R, Pipes GD, Wettenhall REH, Mika H, Hood LE. 1990. Covalent attachment of peptides for high-sensitivity solid-phase sequence analysis. *Anal. Biochem.* 187:56–65
49. Wettenhall REH, Aebersold RH, Hood LE. 1991. Solid-phase sequencing of ^{32}P-labeled phosphopeptides at picomole and subpicomole levels. *Methods Enzymol.* 201:186–99
50. Tempst P, Woo DL, Teplow DB, Aebersold R, Hood L, Kent SBH. 1986. Microscale structure analysis of a high-molecular-weight, hydrophobic membrane glycoprotein fraction with platelet-derived growth factor–dependent kinase activity. *J. Chromatogr.* 359:403–12
51. Aebersold RH, Pipes G, Hood LE, Kent SBH. 1988. N-terminal and internal sequence determination of microgram amounts of proteins separated by isoelectric focusing in immobilized pH gradients. *Electrophoresis* 9:520–30
52. Harrington MG, Gudeman D, Zewert T, Hood L. 1991. Analytical and micropreparative two-dimensional electrophoresis of proteins. *METHODS: Companion Methods Enzymol.* 3:98–108
53. Harrington MG, Hood L, Puckett C. 1991. Simultaneous analysis of phosphoproteins and total cellular proteins from PC12 cells. *METHODS: Companion Methods Enzymol.* 3:135–41
54. Mononen I, Heisterkamp N, Kaartinen V, Williams JC, Yates JR III, et al. 1991. Aspartylglycosaminuria in the Finnish population: identification of two point mutations in the heavy chain of glycoasparaginase. *Proc. Natl. Acad. Sci. USA* 88:2941–45
55. Kaartinen V, Williams J, Tomich J, Yates JR III, Hood L, Mononen I. 1991. Glycosaparaginase from human leukocytes. Inactivation and covalent modification with diazo-oxoriorvaline. *J. Biol. Chem.* 266:5860–69
56. Griffin PR, Coffman JA, Hood LE, Yates JR III. 1991. Structural analysis of proteins by capillary HPLC electrospray tandem mass spectrometry. *Int. J. Mass Spectrom. Ion Process.* 111:131–49
57. Yates JR III, Zhou J, Griffin PR, Hood LE. 1991. Computer aided interpretation of low energy MS/MS mass spectra of peptides. *Tech. Protein Chem. II* 46:477–85
58. Lai E, Davi NA, Hood LE. 1989. Effect of electric field switching on the electrophoretic mobility of single-stranded DNA molecules in polyacrylamide gels. *Electrophoresis* 10:65–67
59. Birren BW, Hood L, Lai E. 1989. Pulsed field gel electrophoresis: studies of DNA migration made with the programmable, autonomously controlled electrode (PACE) apparatus. *Electrophoresis* 10:302–9

60. Landegren U, Kaiser R, Sanders J, Hood L. 1988. A ligase-mediated gene detection technique. *Science* 241:1077–80
61. Nickerson DA, Kaiser R, Lappin S, Stewart J, Hood L, Landegren U. 1990. Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay. *Proc. Natl. Acad. Sci. USA* 87:8923–27
62. Roach JC, Boysen C, Wang K, Hood L. 1995. Pairwise end sequencing: a unified approach to genomic mapping and sequencing. *Genomics* 26:345–53
63. Rowen L, Koop BF, Hood L. 1996. The complete 685-kb DNA sequence of the human β T cell receptor locus. *Science* 272:1755–62
64. Boysen C, Simon MI, Hood L. 1997. Analysis of the 1.1-Mb human α/δ T-cell receptor locus with bacterial artificial chromosome clones. *Genome Res.* 7:330–38
65. MHC Consortium. 1999. Complete sequence and gene map of a human major histocompatibility complex: the MHC sequencing consortium. *Nature* 401:921–23
66. Glusman G, Rowen L, Lee I, Boysen C, Roach JC, et al. 2001. Review: comparative genomics of the human and mouse T-cell receptor loci. *Immunity* 15:337–49
67. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, et al. 1992. Cloning and stable maintenance of 300-kb-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* 89:8794–97
68. Boysen C, Simon MI, Hood L. 1997. Fluorescence-based sequencing directly from bacterial and P1-derived artificial chromosomes. *Benchmarks* 23:978–82
69. Venter JC, Smith HO, Hood L. 1996. A new strategy for genome sequencing. *Nature* 381:364–66
70. Mahairas GG, Wallace JC, Smith K, Swartzell S, Holzman T, et al. 1999. Sequence-tagged connectors: a sequence approach to mapping and scanning the human genome. *Proc. Natl. Acad. Sci. USA* 96:9739–44
71. Hood L. 1992. Biology and medicine in the twenty-first century. In *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, ed. DJ Kevles, L Hood, pp. 136–63. Cambridge, MA: Harvard Univ. Press
72. Koop BF, Hood L. 1994. Striking sequence similarity over almost 100 kilobases of human and mouse T-cell receptor DNA. *Nat. Genet.* 7:48–53
73. Heilig R, Eckenberg R, Petit JL, Fonknechten N, Da Silva C, et al. 2003. The DNA sequence and analysis of human chromosome 14. *Nature* 421:601–7
74. Zody MC, Garber M, Sharpe T, Young SK, Rowen L, et al. 1006. Analysis of the DNA sequence and duplication history of human chromosome 15. *Nature* 440:671–75
75. Int. Hum. Genome Seq. Consort. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:931–45
76. Davidson EH, Rost JP, Oliveri P, Ransick A, Calestani C, et al. 2002. A genomic regulatory network for development. *Science* 295:1669–78
77. Davidson EH, McClay DR, Hood L. 2003. Regulatory gene networks and the properties of the developmental process. *Proc. Natl. Acad. Sci. USA* 100:1475–80
78. Blanchard AP, Kaiser RJ, Hood LE. 1996. High-density oligonucleotide arrays. *Biosens. Bioelectron.* 11:687–90

79. Blanchard AP, Hood L. 1996. Sequence to array: probing the genome's secrets. *Nat. BioTechnol.* 14:1649
80. Lausted C, Dahl T, Warren C, King K, Smith K, et al. 2004. POSaM: a fast, flexible, open source, ink-jet oligonucleotide synthesizer and microarrayer. *Genome Biol.* 5:R58
81. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*. 17:994–99
82. Eng JK, McCormack AL, Yates JR III. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5:976–89
83. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8:175–85
84. Ideker T, Galitski T, Hood L. 2001. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* 2:343–72
85. Hood L, Galas DJ. 2003. The digital code of DNA. *Nature* 421:444–48
86. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, et al. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292:929–34
87. Weston AD, Baliga NS, Bonneau R, Hood L. 2003. Systems approaches applied to the study of *Saccharomyces cerevisiae* and *Halobacterium* sp. *Cold Spring Harbor Symp. Quant. Biol.* 68:345–57
88. Hood L, Galas DJ, Dewey G, Wilson J, Veras R. 2008. *Biology as an Informational Science and the Emergence of Systems Biology*. Greenwood Village, CO: Roberts & Co. In preparation
89. Ramsey SA, Smith JJ, Orrell D, Marelli M, Petersen TW, et al. 2006. Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. *Nat. Genet.* 38:1082–87
90. Gilchrist M, Thorsson V, Li B, Rust AG, Korb M, et al. 2006. Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. *Nature* 441:173–78
91. Taylor RJ, Siegel AF, Galitski T. 2007. Network motif analysis of a multi-mode genetic-interaction network. *Genome Biol.* 8:R160
92. Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, et al. 2007. A predictive model for transcriptional control of physiology in a free living cell. *Cell* 131:1354–65
93. Facciotti MT, Reiss DJ, Pan M, Kaur A, Vuthoori M, et al. 2007. General transcription factor specified global gene regulation in archaea. *Proc. Natl. Acad. Sci. USA* 104:4630–35
94. Nesvizhskii AI, Vitek O, Aebersold R. 2007. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods* 4:787–97
95. Klimek J, Eddes JS, Hohmann L, Jackson J, Peterson A, et al. 2008. The standard protein mix database: a diverse data set to assist in the production of improved peptide and protein identification software tools. *J. Proteome Res.* 7(1):96–103

96. Zhou Y, Aebersold R, Zhang H. 2007. Isolation of N-linked glycopeptides from plasma. *Anal. Chem.* 79:5826–37
97. Mallick P, Schirle M, Chen SS, Flory MR, Lee H, et al. 2007. Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol.* 25:125–31
98. Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, et al. 2006. The inferelator: an algorithm for learning parsimonious regulatory networks from systems biology data sets de novo. *Genome Biol.* 7:R36
99. Shannon PT, Reiss DJ, Bonneau R, Baliga NS. 2006. The Gaggle: an open-source software system for integrating bioinformatics software and data sources. *BMC Bioinformatics* 7:176
100. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13:2498–504
101. Reiss DJ, Baliga NS, Bonneau R. 2006. Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. *BMC Bioinformatics* 7:280
102. Longabaugh WJ, Davidson EH, Bolouri H. 2005. Computational representation of developmental genetic regulatory networks. *Dev. Biol.* 283:1–16
103. Ng WV, Ciufo SA, Smith TM, Bumgarner RE, Baskin D, et al. 1998. Snapshot of a large dynamic replicon in a halophilic archaeon: megaplasmid or minichromosome? *Genome Res.* 8:1131–41
104. Hood L, Heath JR, Phelps ME, Lin B. 2004. Systems biology and new technologies enable predictive and preventative medicine. *Science* 306:640–43
105. Heath JR, Phelps ME, Hood L. 2003. Nanosystems biology. *Mol. Imaging Biol. Official Publ. Acad. Mol. Imaging* 5:312–25
106. Weston AD, Hood L. 2004. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J. Proteome Res.* 3:179–96
107. Lin B, White JT, Lu W, Xie T, Utleg AG, et al. 2005. Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: a systems approach to disease. *Cancer Res.* 65:3081–91
108. Price ND, Edelman LB, Lee I, Yoo H, Hwang D, et al. 2007. Systems biology and the emergence of systems medicine. In *Genomic and Personalized Medicine: From Principles to Practice*, ed. G Ginsburg, H Willard. San Diego: Academic. In press
109. Nelson PR, Goulter AB, Davis RJ. 2005. Effective analysis of genomic data. *Methods Mol. Med.* 104:285–312
110. Sun B, Ranish JA, Utleg AG, White JT, Yan X, et al. 2006. Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics. *Mol. Cell. Proteomics* 6:141–49
111. Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, et al. 2002. Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* 4:323–33

112. Hwang D, Smith JJ, Leslie DM, Weston AD, Rust AG, et al. 2005. A data integration methodology for systems biology: experimental verification. *Proc. Natl. Acad. Sci. USA* 102:17302–7
113. Hwang D, Rust AG, Ramsey S, Smith JJ, Leslie DM, et al. 2005. A data integration methodology for systems biology. *Proc. Natl. Acad. Sci. USA* 102:17296–301
114. Glusman G, Qin S, Raafat El-Gewely M, Siegel AF, Roach JC, et al. 2006. A third approach to gene prediction suggests thousands of additional human transcribed regions. *PLoS Comput. Biol.* 2:e18
115. Price ND, Trent J, El-Naggar AK, Cogdell D, Taylor E, et al. 2007. Highly accurate two-gene classifier for differentiating gastrointestinal stromal tumors and leiomyosarcomas. *Proc. Natl. Acad. Sci. USA* 104:3414–19
116. Huang GM, Farkas J, Hood L. 1996. High-throughput cDNA screening utilizing a low order neural network filter. *BioTechniques* 21:1110–14
117. Smith TM, Abajian C, Hood L. 1997. Hopper: software for automating data tracking and flow in DNA sequencing. *CABIOS* 13:175–82
118. Smith TM, Hood L. 1999. What are biologists going to do with all these data? *Math. Model. Sci. Comput.* 9:155–62
119. Siegel AF, van den Engh G, Hood L, Trask B, Roach JC. 2000. Modeling the feasibility of whole genome shotgun sequencing using a pairwise end strategy. *Genomics* 68:237–46
120. Bonneau R, Baliga NS, Deutsch EW, Shannon P, Hood L. 2004. Comprehensive de novo structure prediction in a systems-biology context for the archaea *Halobacterium* sp. NRC-1. *Genome Biol.* 5:R52
121. Facciotti MT, Bonneau R, Hood L, Baliga NS. 2004. Systems biology experimental design—considerations for building predictive gene regulatory network models for prokaryotic systems. *Curr. Genomics* 5:1389–2029
122. Ideker T, Thorsson V, Seigel AF, Hood L. 2000. Testing for differentially expressed genes by maximum-likelihood analysis of microarray data. *J. Comput. Biol.* 7:805–17
123. Thorsson V, Hörnquist M, Siegel AF, Hood L. 2005. Reverse engineering galactose regulation in yeast through model selection. *Stat. Appl. Genet. Mol. Biol.* 4:article 28
124. Davidson EH, McClay DR, Hood L. 2003. Regulatory gene networks and the properties of the developmental process. *Proc. Natl. Acad. Sci. USA* 100:1475–80
125. McIndoe RA, Stanford JL, Gibbs M, Jarvik GP, Brandzel S, et al. 1997. Linkage analysis of 49 high-risk families does not support a common familial prostate cancer–susceptibility gene at 1q24–25. *Am. J. Hum. Genet.* 61:347–353
126. Jarvik G, Stanford JL, Goode EL, Hood L, Ostrander EA. 1999. Confirmation of prostate cancer susceptibility genes using high-risk families. *J. Natl. Cancer Inst.* 26:81–87
127. Gibbs M, Stanford JL, McIndoe RA, Jarvik GP, Kolb S, et al. 1999. Evidence for a rare prostate cancer–susceptibility locus at chromosome 1p36. *Am. J. Hum. Genet.* 64:776–87

128. Goode E, Stanford JL, Chakrabarti L, Gibbs M, Kolb S, et al. 2000. Linkage analysis of 150 high-risk prostate cancer families at 1q24–25. *Genet. Epidemiol.* 18:251–75
129. Gibbs M, Stanford JL, Jarvik GP, Janer M, Badzioch M, et al. 2000. A genomic scan of families with prostate cancer identifies multiple regions of interest. *Am. J. Hum. Genet.* 67:100–9
130. Janer M, Friedrichsen DM, Stanford JL, Badzioch MD, Kolb S, et al. 2003. Genomic scan of 254 hereditary prostate cancer families. *Prostate* 57:309–19
131. Friedrichsen DM, Stanford JL, Isaacs SD, Janer M, Chang BL, et al. 2004. Identification of a prostate cancer susceptibility locus on chromosome 7q11–21 in Jewish families. *Proc. Natl. Acad. Sci. USA* 101:1939–44
132. Pierce BL, Friedrichsen-Karyadi DM, McIntosh L, Deutsch K, Hood L, et al. 2007. Genomic scan of 12 hereditary prostate cancer families having an occurrence of pancreas cancer. *Prostate* 67:410–15
133. Johanneson B, Deutsch K, McIntosh L, Friedrichsen-Karyadi DM, Janer M, et al. 2007. Suggestive genetic linkage to chromosome 11p11.2-q12.2 in hereditary prostate cancer families with primary kidney cancer. *Prostate* 67:732–42
134. Hood L, Lange PH. 1997. The coming revolution in urology. *Contemp. Urol.* 9:33–50
135. Hood L, Lange PH. 1997. Preparing for the urologic revolution. *Contemp. Urol.* 9:39–58
136. Liu AY, True LD, LaTray L, Nelson PS, Ellis WJ, et al. 1997. Cell-cell interaction in prostate gene regulation and cytodifferentiation. *Proc. Natl. Acad. Sci. USA* 94:10705–10
137. Liu AY, Corey E, Vessella RL, Lange PH, True LD, et al. 1997. Identification of differentially expressed prostate genes: increased expression of transcription factor ETS-2 in prostate cancer. *Prostate* 30:145–53
138. Nelson P, Ng W-L, Schummer M, True LD, Liu AY, et al. 1998. An expressed-sequence-tag database of the human prostate: sequence analysis of 1,168 cDNA clones. *Genomics* 47:12–25
139. Hawkins V, Doll D, Bumgarner R, Smith T, Abajian C, et al. 1999. PEDB: the prostate expression database. *Nucleic Acids Res.* 27:204–8
140. Nelson PS, Hawkins V, Schummer M, Bumgarner R, Ng W, et al. 1999. Negative selection: a method for obtaining low-abundance cDNAs using high-density cDNA clone arrays. *Genet. Anal. Biomol. Eng.* 15:209–15
141. Lin B, White JT, Ferguson C, Bumgarner R, Friedman C, et al. 2000. PART-1: a novel human prostate–specific, androgen-regulated gene that maps to chromosome 5q121. *Cancer Res.* 60:858–63
142. Liu AY, Nelson PS, van den Engh G, Hood L. 2002. Human prostate epithelial cell-type cDNA libraries and prostate expression patterns. *Prostate* 50:92–103
143. Nelson PS, Clegg N, Arnold H, Ferguson C, Bonham M, et al. 2002. The program of androgen-responsive genes in neoplastic prostate epithelium. *Proc. Natl. Acad. Sci. USA* 99:11890–95

144. Lin B, White JT, Lu W, Xie T, Utleg AG, et al. 2005. Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: a systems approach to disease. *Cancer Res.* 65:3081–91
145. Bailey RC, Kwong GA, Radu CG, Witte ON, Heath JR. 2007. DNA-encoded antibody libraries: a unified platform for multiplexed cell sorting and detection of genes and proteins. *J. Am. Chem. Soc.* 129:1959–67