perspectives Journal of proteome research

# A Personal View of Molecular Technology and How It Has Changed Biology

**Leroy Hood**

*President and Director, Institute for Systems Biology, Seattle, Washington 98103*

Received June 27, 2002

I have had the pleasure of practicing biology and technology development for the past 35 years—first at the National Institutes of Health (NIH) for 3 years, then at the California Institute of Technology (Caltech) for 22 years, subsequently at the University of Washington (UW) for 8 years, and finally at the Institute for Systems Biology (ISB) for the past 2 years. In each case, my geographical transitions were prompted by exciting new opportunities and evolving visions of how biology should be practiced. The technology developments I have been associated with over this period helped to catalyze two emerging paradigm changes in biology: systems biology and predictive/preventive medicine.

My career has been driven by two imperatives from my Ph.D. advisor, Bill Dreyer (Caltech). First, "Always practice biology at the leading edge." For much of my career, this leading edge was molecular immunology.[1−14] Second, "With new technologies, biologists have the chance to open up new horizons for exploration in biology." Indeed, Freeman Dyson made this point eloquently (Figure 1). This view is the focus of my discussion.

Several points should be stressed regarding biology and technology. First, technology is all about deciphering biological information, and that information is of three general types: (1) the digital information of DNA; (2) the three-dimensional information of proteins, the molecular machines of life; and (3) the four-dimensional (time-variant) information of biological systems operating across developmental and/or physiological time spans. Second, biology should be the driver for new technologies. Where barriers to deciphering biological information exist, they need to be lifted by appropriate technologies. Third, technology includes new instrumentation (e.g., the automated DNA sequencer) as well as new strategies (e.g., the oligonucleotide ligase assay, or OLA). Finally, challenging new technologies often require the integration of expertise from biology, chemistry, computer science, engineering, mathematics, and physics. Especially important is the development and integration of computational tools for capturing, storing, and analyzing biological information.

Throughout out my career, I have been associated with outstanding colleagues (Table 1), and it is to them that much of the credit must go for our accomplishments. It is with considerable pride that I note many of my former colleagues are today's technology leaders.

Over my career, there have been six successive evolutionary stages to my thinking about biology and technology (Table 2). I will discuss each of these in turn.

"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."

**Figure 1.** Quotes from Freeman J. Dyson, *Imagined Worlds* (Harvard University Press: Cambridge, MA, 1998).

**Table 1.** Colleagues for Instrument and Strategy Development

| | |
|---|---|
| Ruedi Aebersold | protein blotting, protein microsequencing |
| Bruce Birren | pulse-field gel electrophoresis |
| Alan Blanchard | ink-jet DNA synthesizer |
| Ian Clark-Lewis | long peptides |
| Cecilie Boysen | BAC shotgun sequencing |
| Pat Griffin | mass spectrometry and proteins |
| Mike Harrington | two-dimensional gels and proteins |
| Susan Horvath | DNA synthesizer |
| Henry Huang | DNA sequencer (Maxam−Gilbert) |
| Mike Hunkapiller | gas−liquid-phase protein sequencer, DNA synthesizer, DNA sequencer |
| Tim Hunkapiller | computational tools, DNA sequencing |
| Trey Ideker | systems computational tools |
| Karen Jonscher | mass spectrometry and proteins |
| Rob Kaiser | labeling DNA |
| Steve Kent | protein synthesizer, long peptides |
| Joan Kobori | primer-directed sequencing |
| Eric Lai | pulse field gel electrophoresis |
| Steve Lasky | ink-jet synthesizer |
| Ulf Landegren | oligonucleotide ligase assay (OLA)/SNP analyses |
| Greg Mahairas | BAC-end sequence mapping, or sequence tagged connector (STC) 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 |
| Paul Tempst | protein microsequencing |
| David Teplow | protein chemistry |
| Mike Waterfield | solid-phase protein sequencing |
| John Yates | mass spectrometry and proteins |

## New Technologies Open Up New Horizons of Biology for Exploration

I moved from NIH to Caltech in 1970 because I had decided that I would like to divide my time between technology and biology. Caltech seemed an ideal environment for both biology and technology development. During my time at Caltech, my

10.1021/pr020299f CCC: $22.00 © xxxx American Chemical Society 
PAGE EST: 10.5
Published on Web 00/00/0000

Table 2. Deciphering Biological Information: Principles

- new technologies open up new horizons of biology for exploration principles
- integration of technologies creates new analytic opportunities
- the Human Genome Project introduces discovery science, which opens the door to global analyses
- cross-disciplinary environments are essential for applying engineering and computational tools to biology
- systems biology is the integration of biology, medicine, computation, and technology to decipher and model biological information
- predictive and preventive medicine will transform the practice of medicine

Table 3. Instruments Developed by the Hood Laboratory

| instrument | special features | colleagues |
|---|---|---|
| gas–liquid-phase protein sequencer | 100-fold more sensitivity | M. Hunkapiller |
| DNA synthesizer | robust, high repetitive yields | S. Horvath, M. Hunkapiller |
| peptide synthesizer | high repetitive yield | S. Kent, Applied Biosystems |
| DNA sequencer | automated DNA sequencing | L. Smith, M. Hunkapiller |
| ink-jet oligonucleotide synthesizer | synthesize oligonucleotide arrays | A. Blanchard |

group evolved a then unique cross-disciplinary culture where biologists and technologists interacted freely and effectively. Over the next 22 years, we developed prototypes for four instruments and began thinking about a fifth (Table 3). Each instrument met the criterion that it opened up new horizons of biology for exploration. Let me discuss each of these instruments in turn.

**Protein Sequencer.** I was initially a protein chemist; hence, my beginning technology efforts focused on highly sensitive protein sequencing. The protein sequencer, developed by Mike Hunkapiller, used the Edman chemistry to sequence proteins from their N-terminus. This gas–liquid-phase protein sequencer was ~100 times more sensitive than its contemporary counterparts.[15,16] This 100-fold increase in sensitivity allowed us to sequence a series of fascinating proteins—available in low quantities and, hence, heretofore inaccessible to sequence analysis. Strikingly, the sequence analyses of six of these proteins opened up six new areas in biology.

(1) Platelet-derived growth factor is a blood hormone whose N-terminal sequence revealed a near sequence identity to an avian oncogene, *v-cis*.[17] We postulated, for the first time, that oncogenes might be ordinary cellular genes, for example, controlling growth, DNA repair, or the cell cycle that have been captured by viruses and subjected to viral control. This was a major step forward in thinking about oncogenes. In addition, this analysis was the first application of bioinformatics to biology; that is, comparing a newly determined protein sequence against a database of all known preexisting sequences to find interesting homology relationships (Mark Boguski, personal communication).

(2) Together with Amgen, one of the first biotechnology companies, we determined the N-terminal sequence of erythropoietin, a blood hormone that stimulates the growth and development of red blood cells. Amgen used this sequence information to clone the corresponding gene. Erythropoietin, biotechnology's first billion dollar drug, was extremely effective in dealing with chronic anemia.

(3) Together with Dr. Stanley Prusiner, we determined the first sequence of the prion protein.[18] This protein turned out to be an extremely difficult sequence, as it was purified by proteolytic digestion and, hence, had several N-terminal sequences (e.g., a ragged N-terminus) which needed to be placed in phase with one another. This protein sequence was used to synthesize a DNA probe and clone the prion gene.[19] These studies led to the fascinating observation that prion diseases (e.g., mad cow disease and human Cruetzfild–Jacob disease) are initiated by misfolded prions that are pathogenic and have the ability to catalyze the misfolding of normal prions, a new mechanism for disease. This work led to a Nobel Prize for Stanley Prusiner.

(4) We were the first to sequence the human $\alpha$ and $\beta$ interferons.[20,21] These sequences led to the cloning of the corresponding genes and opened up our ability to understand how interferons stimulate T-cell effector functions. Some of the interferons have proved to be effective drugs for dealing with certain kinds of cancers.

(5) We were the first to sequence a colony-stimulating factor, key hormones for the development of hematopoietic lineages. The sequence of the mouse macrophage-granulocyte colony-stimulating factor (GMCSF) led to the cloning of the mouse and then human GMCSFs. These molecules provided new insights into the development of one branch of the hematopoietic cell lineage. The human GMCSF has been a useful drug for certain diseases.

(6) Together with Mike Raftery (then at Caltech), we determined the N-terminal sequences for the four chains of the acetylcholine receptor in the *Torpedo calcifornica*.[22] Professor Numa at Kyoto University then used these sequences to clone the corresponding genes and started his now classical studies of receptors of the nervous system.

So the highly sensitive protein sequencer opened up a multiplicity of new areas in biology through the sequence analyses of heretofore inaccessible proteins and the cloning of their corresponding genes. Truly, this technology found new things that needed to be explained (Figure 1).

**Protein Synthesizer.** Steve Kent, a former student of Robert Merrifield, the developer of solid-phase peptide synthesis, joined our laboratory to optimize the peptide synthesis chemistry so that very long peptides could be synthesized through high repetitive yields. Steve, together with Applied Biosystems, developed an instrument for this optimized chemistry.[23] This instrument allowed us to synthesize, for example, the 140 residue IL3 molecule and highly purify it to carry out fascinating structure/function studies.[24] Steve Kent's most striking accomplishment was the synthesis, purification to homogeneity, and crystallization of the 99-residue HIV protease. Together with the drug company Merck, Steve determined the structure of this molecule—a key step in generating one of the most effective of all AIDS therapeutic drugs, the protease inhibitor. Finally, we went on to synthesize a series of Zinc finger domains to decipher their strategies for binding, specifically to unique DNA sequences.[25] The highly efficient peptide synthesizer opened up many new directions for the study of proteins and peptides.

**DNA Synthesizer**. Susan Horvath, Mike Hunkapiller, and I developed a robust DNA synthesizer.[26] I remember going to

Table 4. Strategies for Genomic Analysis Developed by the Hood Laboratory

| technique | Hood laboratory scientist | ref |
|---|---|---|
| 1. primer-directed sequencing | J. Kobori | 30 |
| 2. oligonucleotide ligase array for SNPs | U. Landegren | 31 |
| 3. shotgun sequencing for large genomic inserts | L. Rowen, C. Boysen | 32 |
| 4. sequence-tagged sites (STSs) for physical mapping | L. Hood | 33 |
| 5. sequence-tagged connectors (STCs) for virtual physical maps and genome scans | G. Mahairas | 34, 35 |

Marvin Caruthers, inventor of the phosphoamidite DNA synthesis chemistry, early in the 1980s and arguing that we would like to learn the DNA synthesis chemistry so as to automate it. Coruthers was skeptical about the need for automation, arguing that the demand would never be great and that he could teach virtually anyone how to synthesize DNA in a week. We did develop the DNA synthesizer. It provided a variety of key reagents for molecular biology, including DNA probes for primer-directed DNA sequencing and DNA mapping, probes for cloning genes, and even the overlapping DNA fragments necessary to synthesize complete genes. The ready availability of DNA fragments certainly catalyzed many new strategies in molecular biology. Indeed, if DNA primers were not easy to synthesize, the polymerase chain reaction might never have been invented. Thus, one technology may enable the creation (and development) of other technologies.

**DNA Sequencing.** In my laboratory, Henry Huang started in 1979 to automate the Maxam–Gilbert sequencing procedure. These efforts were unsuccessful, but they did create a foundation for our next efforts. Then Lloyd Smith, Tim Hunkapiller, and I came up with the idea of using four fluorescent probes and capillary electrophoresis to automate the Sanger DNA sequencing procedure. Lloyd Smith went on to build a prototype instrument in conjunction with Mike Hunkapiller and his colleagues at Applied Biosystems.[27] It is interesting to note that two attempts in the early 1980s to obtain NIH support for automated sequencing failed. The reviewers argued that "It was impossible" or that "Graduate students could do it less expensively." These responses illustrate a limitation of the otherwise outstanding NIH peer review system—it is often incapable of dealing effectively with dramatically new opportunities in technology (or biology). The automated fluorescent DNA sequencer has been the key technology for sequencing of the genomes of the human, a variety of model organisms and many microbes. The DNA sequencer ushered in the post-genome era in biology.

**Ink-Jet Oligonucleotide Arrays.** In the early 1990s, Alan Blanchard began to apply the delivery technology of ink-jet printers to the synthesis of oligonucleotide arrays on glass.[28,29] Our current version of the ink-jet DNA synthesizer has 192 piezoelectric pumps that each deliver 14 000 6 pL droplets/s. Thus, the individual base components of thousands of oligonucleotides can rapidly and specifically be sprayed across a glass slide under computer control. Affymetrix had pioneered the synthesis of oligonucleotide arrays using photolithographic techniques, a process that had two striking advantages: density (>100 000 oligonucleotides per array) and mass production. Our ink-jet technology also had striking advantages: flexibility in format design (controlled by a computer and not expensive photolithographic masks) and a high repetitive synthesis yield, meaning that long oligonucleotides (up to 60 bases) could be synthesized. In the last year, Steve Lasky has applied the ink-jet technology to the synthesis of double-stranded oligonucleotide arrays (for assaying transcription factor binding sites). More recently, we have talked with Steve Kent about the possibility of designing protocols for synthesizing arrays of different protein domains (~100 residues). We have also used the ink-jet oligonucleotides to interrogate patterns of gene expression and single nucleotide polymorphisms (SNPs). The ink-jet technology was initially licensed to Rosetta Bioinformatics and more recently to Agilent. Certainly, DNA (oligonucleotide) arrays have transformed contemporary biology.

**Strategies for Genome Analysis.** I would emphasize that molecular strategies for analyses are as important as new instrumentation. In Table 4 are given a list of strategies my laboratory developed for genomic analyses. Let us consider two of these strategies, both of which opened up striking new opportunities for genomic analyses. First, Ulf Landegren developed the oligonucleotide ligase assay, which allowed us to analyze single nucleotide polymorphisms with an assay that could be automated for very high throughputs.[31] We synthesized three 25-mer oligonucleotides (Figure 2)—the 5′ oligonucleotide had a 5′ biotin tag (a hook) and was complementary to the DNA sequence immediately 5′ to the polymorphic site. The two 3′ probes were identical and complementary to sequence 3′ to the polymorphic site but for their 5′ ends at the site of polymorphism where each had a nucleotide complementary to one of the two alleles. The two 3′ probes were labeled with a fluorescent group. The 5′ probe, together with separate 3′ probes (two different reaction mixtures), were added to individual genomic DNA and DNA ligase was added. Only 3′ nucleotides with perfect complementarity to the genomic DNA at their 5′ ends could be joined to their adjacent 5′ counterpart by DNA ligase. Avidin columns would pull out the 5′ nucleotides and, if they were ligated to one (homozygous) or both (heterozygous) 3′ oligonucleotides, the fluorescent tags could identify the SNP. We used this technology to promote the idea of genome-wide SNP analyses long before it assumed its current popularity.

Second, the sequence tagged connector (STC) strategy has provided a powerful new strategy for mapping and sequencing genomes.[34,35] This approach was based on our earlier theoretical analyses of optimal strategies through paired-end insert sequencing for analyzing the human genome.[35–37] The idea is to sequence both ends of 150 000 randomly generated bacterial artificial chromosome (BAC) inserts (average 200 kilobases (kb) in length) from the human genome. This would give a 500 base pair (bp) tag about every 10 kb across the genome. Furthermore, this would provide a random genomic scan of $1.5 \times 10^8$ bp (300 000 STCs $\times$ 500 bp) or a 0.2-fold coverage of the genome. This genomic scan allows biologists to sample genes and repeat sequences across the genome. Finally, the STC approach created a virtual physical genome map that can be realized by sequencing overlapping BAC inserts (Figure 3). Once a 200 kb BAC is sequenced, on average, it will reveal 20 overlapping BACs (one every 10 kb). The BAC-end sequence tags connect the overlapping BAC clones to the sequenced seed BAC and hence were designated sequence-tagged connectors

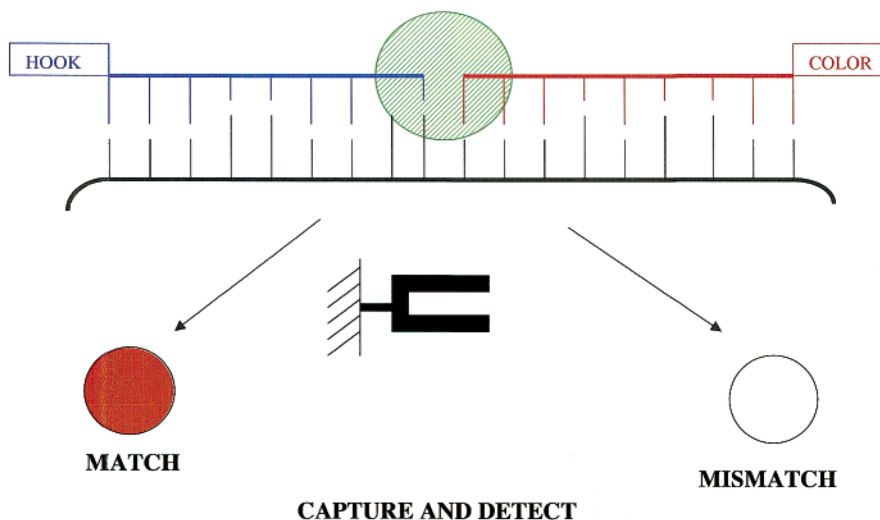


**Figure 2.** Schematic diagram of the oligonucleotide assay: blue, 5′ oligonucleotide probe; red, 3′ oligonucleotide probe; hook, biotin tag; green circle, DNA ligase; black, avidin column. Match means ligase joined the 5′ and 3′ probes because of perfect complementarity at the SNP site. Mismatch means the 5′ and 3′ probes were not joined because of a base mismatch leading to a failure of molecular complementarity (see text).

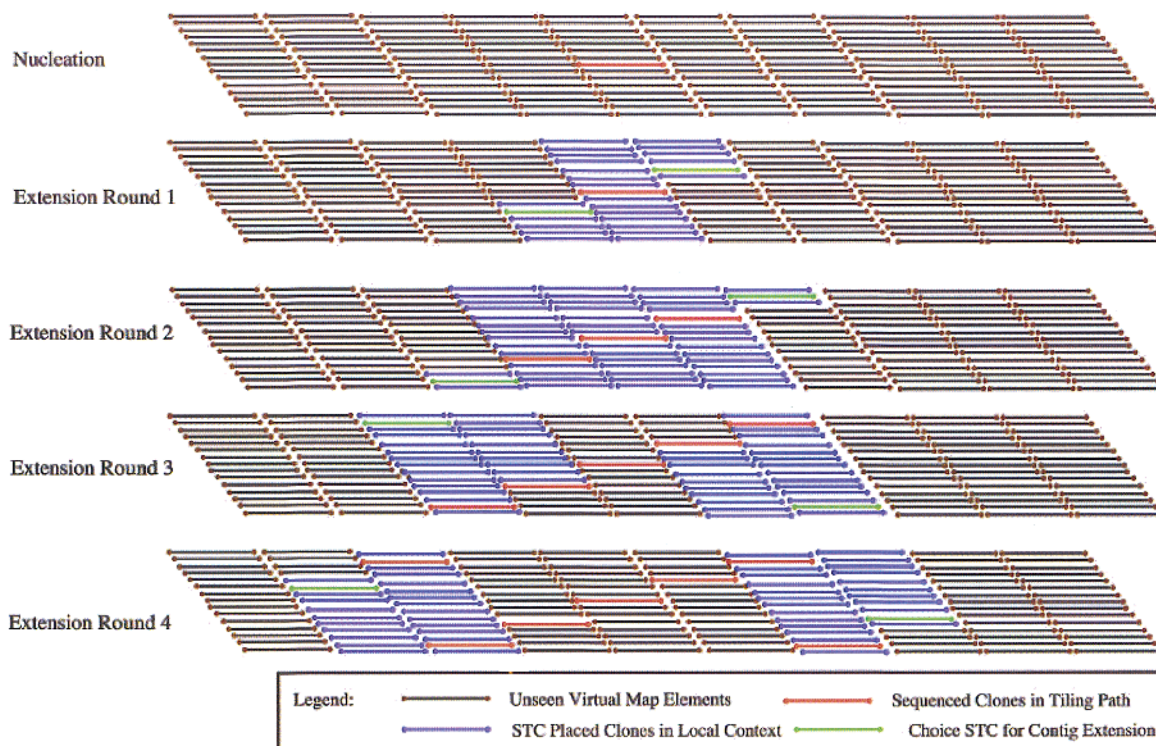


**Figure 3.** Schematic diagram of the virtual physical map created by the sequence tag connector strategy and the process of sequence extension (see text).

(STCs). The BAC clones minimally overlapping on the 5′ and 3′ ends of the sequenced seed BAC clone can be sequenced, and through this iterative process, sequencing can continue until the end of the chromosome is reached (or a clone gap in the overlapping BACs blocks progress). With multiple seed or initial BAC sequences, complete genome sequencing can start from many sites. This is an enormously efficient approach to sequencing genomes. Greg Mahairas sequenced the first draft of the rice genome using the STC approach[35] as a collaboration with Monsanto.

Each of the other strategies in Table 4 has also opened up a myriad of additional opportunities for genomic analyses.

**Strategies for Protein Analysis.** Over the years, in addition to developing the highly sensitive protein sequencer, my laboratory developed a variety of new strategies for protein analysis focused on issues of how to increase the sensitivity of protein sequencing and an exploration of the mass spectrometer as a global tool for analyzing proteins. Jack Silver explored the use of radioactive phenyl isothiocyanate for protein microsequencing.[39] Mike Hunkapiller explored new spinning cup

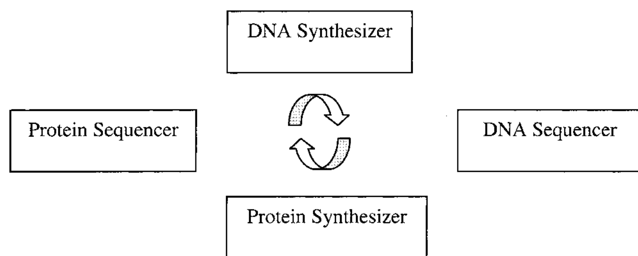


**Figure 4.** Integration of technologies: the microchemical facility (see text).

carriers[40] and acrylamide approaches to the preparation of proteins for microsequencing.[41] Ruedi Aebersold introduced protein electroblotting,[42,43] the use of nitrocellulose to separate peptides cleaved after blotting for internal sequencing,[44] and solid-phase covalent attachment for the microsequencing of peptides[45] and phosphopeptides.[46] John Yates helped pioneer the application of mass spectrometry to the sequence analysis of proteins,[47,48] was one of the first to apply electrospray techniques to proteins,[49] and started his very important work on computer-aided global analyses of peptide mass spectra.[50]

## Integration of Technologies Creates New Analytic Opportunities

**Microchemical Facility.** The idea of integrating the power of the protein and DNA synthesizers and sequencers to decipher biological information more effectively emerged congruently with our vision to develop them (Figure 4). We submitted a paper to *Nature* on this integrated microchemical facility, which took almost two years for publication because of skepticism concerning some of the claims.[51] Let me illustrate the use of the integrated microchemical facility by describing our approach to the prion problem. We began by sequencing the N-terminus of the prion protein (protein sequencer). Next, we synthesized a degenerate DNA probe after translating the protein sequence into DNA sequence (**DNA synthesizer**) and used this probe to clone a full-length cDNA and several genomic clones containing the gene. The cDNA and genomic clones were then sequenced (**DNA sequencer**). The sequence of the cDNA clone suggested candidate regions as possible nuclei for protein misfolding. The sequence of the genomic clone revealed a second closely-linked prion-like gene termed dopple (second) that has fascinating biological behavior.[52] Finally, the prion gene was translated into a protein sequence, and peptide fragments were synthesized (**peptide synthesizer**) and used to raise specific antibody. Accordingly, the integrated microchemical facility allowed us to move from protein sequence to gene sequence and, finally, to biology with antibodies to characterize prion protein behavior. These approaches were complemented by others from classic molecular biology (knockouts, expression vectors, etc.).

**High-Throughput Platforms.** Technologies can be integrated in a second way—to develop high-throughput platforms for the rapid analysis of biological information. For example, driven by the Human Genome Project, a factory-like production line for high-throughput DNA sequencing has been developed with three major components: a front-end to automatically produce DNA fragments for sequencing (robots and automated liquid handling); a 96-capillary DNA sequencer with automated sample injection (now capable of sequencing more than 1.5 million bases/day); and computational tools for the capture, quality assessment, storage, and distribution of the sequence information. Indeed, my group pioneered the development of two of the major instruments for high-throughput genomic analyses (DNA sequencer, oligonucleotide arrays) and pushed applications of the mass spectrometer for proteomics. We at the Institute for Systems Biology and others are developing high-throughput platforms for SNP analysis, gene expression, and various aspects of proteomics. These tools will allow the hundreds of genomes to be sequenced soon to be analyzed with respect to their polymorphisms, gene expression patterns, and protein expression patterns.

**Three Stages of Instrument Development.** Biological instruments focused on the large-scale acquisition of biological information go through three distinct developmental stages. This process can be illustrated with the DNA sequencer. **(i) The Prototype Stage.** We started to automate DNA sequencing in 1979, and it was 1985−1986 before we had a prototype instrument that clearly demonstrated the feasibility of all the aspects of automated fluorescent DNA sequencing.[27] **(ii) Robust Instrument.** ABI took about 3 years and $75 million to develop the first robust DNA sequencer that could be used routinely in most laboratories. **(iii) Factory-like High-Throughput Production.** Driven by the Human Genome Project, high- throughput production platforms for large-scale DNA sequencing were developed in the late 1990s. In the past, academics have generally confined their efforts to the first of these stages. Industry has carried out the second and third stages because of the cost and engineering required. At the Institute for Systems Biology, we are actively collaborating with industry on the second and third stages of development for one high-throughput proteomics platform.

**Commercialization.** In the late 1970s, a friend told me I should think about commercializing the first of these instruments, the protein sequencer. I liked the idea and went to Caltech's President, Marvin Goldberger, with the idea. President Goldberger said, "Caltech has no interest in commercializing instruments. You will have to do it yourself—if you are really interested." I decided to try. I went to 19 different instrument companies with the fully developed protein sequencer and a vision of how the other three instruments we were or would be developing could change biology. None of the companies were interested. But all was not lost. A San Francisco venture capitalist, Bill Bowes, called me and said, "I hear you've been shopping some instruments around and failed to create any interest. How about the possibility of my providing $2 million to start a company to manufacture these instruments?" I was ecstatic. President Goldberger was less than enthusiastic but he was gradually convinced, as there were no other options. Then I gave a talk on my vision of this instrumentation and its impact on biology to the Caltech trustees. Arnold Beckman came up immediately after my talk and said, "This is wonderful. I want Beckman Instruments to partner with you." In fact, I had visited Beckman Instruments three different times, and the last time I was unequivocally told there was no interest in my vision. After some stumbling around, Caltech agreed to start Applied Biosystems with the venture capital support. (I learned from that experience to pitch visions to CEOs and not middlemen interested primarily in profit and loss.) Applied Biosystems (ABI) today is the premier company in the world for molecular instrumentation, thanks in no small measure to Mike Hunkapiller, who went to ABI and then rose to become its President. In retrospect, it was fortunate that none of the 19 preexisting companies took up the offer to develop the four instruments,

for none of them would have had the total commitment to developing these technologies exhibited by ABI, nor would (could) they have recruited all the diverse talent necessary to effectively develop these instruments. I believe emerging technologies often need new companies to catalyze their successful development.

I have always been committed to transferring knowledge to society and, over the years, have played roles in founding or co-founding 11 different companies, including Applied Biosystems, Amgen, Systemix, Darwin, Rosetta, and MacroGenics—all oriented in different directions. Generally, each employed, at least in part, strategies or technologies spun off from our laboratory. For example, just as the DNA and protein synthesizers and sequencers were the cornerstones of ABI, so too was the ink-jet oligonucleotide synthesizer the foundation of Rosetta.

My interest in automated DNA sequencing then led in a new direction—the Human Genome Project.

## The Human Genome Project Introduces Discovery Science, Which Opens the Door to Global Analyses

In the spring of 1985, I was invited to Santa Cruz, CA, for the first meeting ever held on the Human Genome Project. The Chancellor, Robert Sinsheimer, invited 12 scientists, Wally Gilbert, Charles Cantor, George Church—indeed, most of the principal advocates of the Human Genome Project—to debate the feasibility and merits of sequencing the human genome. After $1^1/_2$ days of fascinating discussion, we reached two conclusions: (1) although difficult, it would be feasible to sequence the human genome, and (2) the group was split 6:6 on whether it was a good idea.

I came away from the meeting with two major impressions. First, the project would introduce to biology a new type of science, which I termed **discovery science**. Discovery science takes a biological object (the genome), defines all of its elements (the sequences of the chromosomes), and places them in a database—to enrich the infrastructure of biology and strengthen the approaches employed by the more classical hypothesis-driven science. The power of discovery science is that it makes comprehensive or global studies possible. For example, by defining all the genes in the human genome, oligonucleotide arrays can interrogate the behaviors of all genes in a normal cell and its cancer counterpart. Second, clearly high-throughput DNA sequencing needed to be developed if the Human Genome Project was ever to be completed. It seemed likely that large-scale DNA sequencing would also drive high-throughput instrumentation for analyzing gene products such as mRNAs and proteins. This was the major rationale for our developing the ink-jet oligonucleotide synthesizer.

The opportunity for high-throughput platforms to decipher biological instrumentation also raised an issue of extreme importance—these high-throughput platforms would have to be developed in a cross-disciplinary environment with biologists, chemists, computer scientists, engineers, mathematicians, and physicists all pooling their collective expertises.

## Cross-Disciplinary Environments Are Essential for Applying Engineering and Computational Tools to Biology

In 1987, the National Science Foundation (NSF) had the first competition for Science and Technology Centers (STCs), and we were awarded an STC for Molecular Biotechnology. Interestingly, only one of the other Biology faculty members at Caltech was interested in participating in the STC (Eric Davidson). The mission of the STC was as follows: (1) to integrate science and technology; (2) to promote meaningful academic/industrial partnerships; and (3) to stimulate K−12 science education (another way of transferring knowledge to society). I believe we were superbly successful in each of these objectives. The STCs had significant (and flexible) budgets (ours was $3 million per year), and they were awarded for 11 years (with two competitive renewals).

**Table 5.** Faculty of the Department of Molecular Biotechnology

| faculty member | cross-disciplinary exptertise |
|---|---|
| Ruedi Aebersold | protein chemistry, proteomics |
| Ger van den Engh | biophysics, cell sorting |
| Phil Green | mathematics, computational biology |
| Joan Goverman | immunology |
| Leroy Hood | immunology, biotechnology, genomics |
| Debbie Nickerson | DNA polymorphisms, human genetic disease |
| Maynard Olson | chemistry, genetics, genomics |
| Barbara Trask | genomics, cell sorting |
| John Yates | mass spectrometry, proteomics |

The STC gave us the resources to become far more cross-disciplinary at the level of my research group; however, the Division of Biology at Caltech was still rather narrowly focused in its commitment to classical molecular, developmental, and neurological biology. I proposed a new cross-disciplinary branch for the Division (Molecular Biotechnology), but the leaders in Biology felt that cross-disciplinary approaches to biology should be carried out in engineering. I then proposed to the President of Caltech, Tom Everhart, that I start a new Division of cross-disciplinary biology. He said, “Fine, if you can convince your biology colleagues.” The Divisions of Engineering and Chemistry were sympathetic to this idea, but Biology could not be persuaded. To be perfectly honest, my group had become large in its efforts to (1) bring in the cross-disciplinary talent I needed from Engineering, Chemistry, and Computer Science (junior colleagues), (2) create a microchemical facility that served more than 40 faculty, and (3) have leading-edge biology to interface with the technology. Biology had traditionally had smaller groups; hence, my group’s size was a concern for some faculty.

I really wanted to create a more effective cross-disciplinary environment with other interdisciplinary faculty participating. This prompted my second academic move—to the University of Washington School of Medicine where a forward-looking Dean of Medicine, Philip Fialkow, agreed that I should start a new cross-disciplinary department, which I named the Department of Molecular Biotechnology. Bill Gates provided $12 million to start the new department, and I moved the STC from Caltech to the University of Washington. The annual STC funding contributed significantly to the department’s highly successful start. I was able to rapidly recruit a strikingly cross-disciplinary faculty, some of whom went on to remarkable accomplishments (Table 5). The following are some examples:

• John Yates and Ruedi Aebersold really started proteomics—John with Sequest, a computer program that allowed mixtures of proteins (peptides) to rapidly be characterized from complex mixtures, and Ruedi with the isotopic-coded affinity tag (ICAT) methodology permitting the quantification of global protein sets in two different cellular states.

• Maynard Olson and I each independently had Genome Centers that contributed to the sequencing of the human genome[53,54] and to diverse strategies and technologies for

genomics (e.g., the ink-jet synthesis technology for oligonucleotide arrays).

• Ger van den Engh developed a multiparameter, high-speed cell sorter that had the highest throughput of any cell sorter in the world.

• Phil Green developed the software programs Phred and Phrap that were used by the entire human genome community to call DNA bases and statistically assess their accuracy.

• Debbie Nickerson carried out beautiful SNP studies in humans.

Thus, the Department of Molecular Biotechnology, the first truly cross-disciplinary department in biology, more than fulfilled my initial expectations. Its scientific growth and evolution continued so that by 1995 a striking new opportunity was emerging—**systems biology**. Systems biology is the study of all the elements in a biological system (all genes, mRNAs, proteins, etc.) and their relationships one to another in response to perturbations—a contrast to the one gene or one protein at a time studies so successful in biology over the preceding 30 years or so.

In 1996, I went to the President of the University of Washington, Richard McCormick, with the proposition I raise the money to fund a new building to house my growing department and develop opportunities for systems biology. McCormick said there were 10 building requests ahead of mine and that it would be 6−10 years or more before I could hope for a building. I then spent the next three and a half years attempting to persuade the School of Medicine that it should have a relatively free-standing Institute for Systems Biology. I failed in this endeavor because the academic administrative structure of a state-funded university was not set up to accommodate the striking integrative organizational changes required by systems biology (see below). In December 1999, I resigned from the University of Washington to co-found with Alan Aderem and Ruedi Aebersold the Institute for Systems Biology, an independent, nonprofit research institute. I was convinced that systems biology will be a central driving force in biology and medicine for the 21st century.

## Systems Biology: The Integration of Biology, Medicine, Computation, and Technology To Decipher and Model Biological Systems

Three advances drove the emergence of systems biology: (1) the Internet made it possible to capture, store, and transmit massive data sets and to communicate effectively on a global scale; (2) a cross-disciplinary environment was essential for the emergence of systems biology and its necessary integration of technology, computation, and biology; and (3) the Human Genome Project has transformed our views and the practice of biology. Let us discuss seven ways the Human Genome Project has impacted biology.

**Biology Is an Informational Science.** This is the most important conceptual change in how we view biology. Three points are worth stressing. (1) The core of biology, the genome, is digital. Biology stands in contrast to all other scientific disciplines—physics, chemistry, geology, etc.—where scientists come to know these disciplines through analog measurements. The important philosophical point is that digital codes can be completely determined. Thus, the core of biology is ultimately knowable, and hence, we start with a certainty that is not possible in the other disciplines. The challenge of biology, of course, is to understand how this digital core along with environmental information is transformed by many analog processes into a four-dimensional (time-variant) organism. (2) The digital core of life (genome) has two different major types of information: the genes, which encode the molecular machines of life, and the regulatory networks that control the behaviors of the genes. All metazoan organs share very similar toolboxes of proteins, although in more complex organisms some gene families or proteins become more complex. What makes metazoan organisms truly different from one another are the regulatory networks (the interactions of individual transcription factors with their complementary transcription factor binding sites). The important point is that the information of regulatory networks is ultimately digital because it is the organization of transcription factor binding sites adjacent to each gene that determines its behavior. Regulatory networks are the key to understanding evolution, development, and even physiology. (3) Biological information is hierarchical and moves through many different levels: DNA → RNA → protein → protein interactions → biomodules → etc. (Figure 5). Biomodules are collections of proteins that execute a particular phenotypic task (e.g., metabolize galactose). The important point is that each level of information provides new insights into the operations of biological systems. Hence, for systems biology, information must be gathered from as many of these levels as possible and integrated.

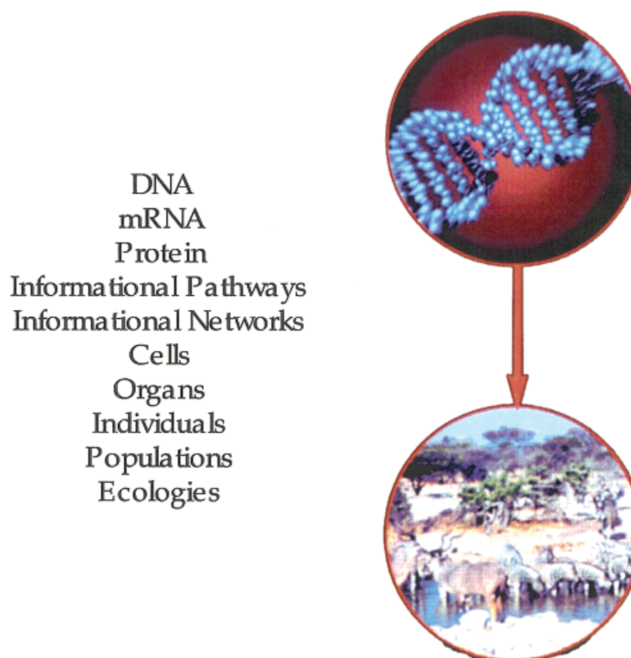


**Figure 5.** Hierarchical levels of biological information.

**Discovery Science Enhances Global Analyses.** Discovery science, by defining all the elements of various systems (e.g., the sequence of the genome, all the mRNAs present in a particular cell or organism—the transcriptome, all the proteins present in a particular cell or organism—the proteome, etc.), opens the possibility of global analyses for the elements of biological systems to be studied.

**A Genetics Parts List Provides a Toolbox of Genetic Elements for Systems Analyses.** The genome of an organism allows one to identify all (most) of its genes and the protein domains encoded by the genes. The identification of the 100 or more protein domains is important because they provide insights into the functioning of the corresponding proteins. The genome sequence provides all of the regulatory sequences

associated with genes and is a major step toward understanding the functioning of regulatory networks (e.g., the identification of the transcription factor binding sites, their transcription factor ligands, and insights into how both interact to trigger the ensuing regulatory behavior). Finally, the human genome provides access to the polymorphisms of humans and opens the possibility of correlating some of these with physiological variations and predispositions to disease.

**High-Throughput Platforms Permit One To Carry Out Global Analyses at the DNA, RNA, and Protein Levels.** Large-scale DNA sequencing, genotyping, DNA (oligonucleotide) arrays, proteomic analyses, and multiparameter high-speed cell sorting are examples of high-throughput platforms capable of capturing enormous amounts of biological information—a central feature of systems biology. The keys for the future in high-throughput platforms are miniaturization, parallelization (many analytic channels), automation, and integration. These requirements are pushing us toward microfluidics and micro-electronics and, ultimately, toward nanotechnology. The critical objectives are higher throughput, higher quality data, and lower cost per unit. Ultimately, we must move toward single-cell and even single-molecule analyses.

To provide a concrete example of a high-throughput platform, consider the DNA sequencer. From our original prototype in 1986 to the most recent 96-capillary ABI sequencers, there has been almost a 3000-fold increase in throughput. My prediction is with single DNA molecule sequencing over the next 10 years there will be another 3000-fold increase in throughput and striking decreases in cost (I predict in 10 years we could sequence a human genome in a day for less than $10,000—in contrast to the more than $50 million to $100 million it would cost today). This means that the digital DNA cores, the genome sequences, of any living organism of interest will be readily accessible.

**Computational, Mathematical, and Statistical Tools Are Essential for Handling the Explosion of Biological Information.** Biologists need tools for capturing, storing, analyzing, assessing quality, graphically displaying, integrating, modeling, and distributing biological information. These represent some of the largest challenges for systems biology. A critical point is that the computer scientists, mathematicians, and statisticians must have a deep understanding of biology if they are to contribute significantly, as must biologists acquire an understanding of these disciplines outside of biology. This cross-disciplinary education is an enormous challenge.

**Model Organisms are Rosetta Stones for Deciphering Biological Information.** Since the genome sequences are available for most model organisms (e.g., *Escherichia coli*, yeast, fly, nematode, mouse, etc.), their genetics parts lists permit biological systems to be perturbed genetically (e.g., knockouts, overexpression) or environmentally to determine the behavior of the systems elements with respect to one another. Moreover, since all organisms share many basic biological systems, model organism studies provide insights into human biological systems. Model organisms will be the Rosetta Stones for deciphering human biological complexity.

**Comparative Genomics Is a Key to Deciphering Biological Complexity.** Computational tools will, in time, allow us to extract all genes and all regulatory networks from the digital genome of any organism, thus providing after integration of these two types of biological information, the logic of life for that organism. Comparing the genomes of different organisms will allow us to ascertain how the logic of life has evolved. These comparisons will be particularly valuable in deciphering the deep logic of regulatory networks.

Each of these seven changes has catalyzed the emergence of systems biology.

## Systems Biology: The Rationale and Implementation

**Rationale.** Biologists have for the past 30 years tended to study biological systems one gene (or one protein) at a time. Systems approaches attempt to study the behavior of all of the elements in a system and relate these behaviors to the systems or emergent properties. For example, the human immune system is composed of $10^{12}$ lymphocytes that exhibit two systems properties: immunity and tolerance (the inability to react against self). In the past, immunologists have studied immune response one gene (or protein) or one or a few cell types at a time. Immunologists, after >30 years of study, do not understand the systems properties of the immune system. Systems approaches will bring fundamental insights into these emergent properties.

As a hypothetical example of systems biology, suppose we consider how a car functions. Biologists traditionally would each study and become experts in one component of the car. Thus, the wheel biologists would study wheels and talk to one another, but they might only rarely talk to the transmission biologists. No one biologist would have even a vague overall picture (understanding) of the car. The systems approach with high-throughput technologies, in contrast, allow us to define all of the elements in the system (car) and measure their relationships one to another as the car is perturbed to carry out its function (before or after the modification of individual elements). For the car, the different elements or data types would be mechanical, electrical, and control (computer chips regulating various functions). Under the systems approach, these diverse types of data would be gathered, integrated, graphically displayed, and ultimately mathematically modeled. The perturbation process would have to be iterated repeatedly, each time with refinement of the model, until the model reflected accurately the systems properties (behavior of the biological information). The ultimate objective would be to accurately predict the behavior of the system given particular perturbations. Indeed, the algorithmic approach to systems biology might be stated as follows:

- Select the biological system and model organism for study.
- Employ discovery science to define the genetic elements of the organism (e.g., its genome sequence, its genes, and proteins).
- Utilize all previous biological knowledge to identify as many elements as possible in the system and, if feasible, generate a preliminary model of how the system works based on previous biology.
- Perturb the digital code by genetic (destroy or enhance the functions of genes in the system) or environmental perturbations of the system and then follow the global behavior of the system's elements throughout development or the activation of physiological responses (this includes measuring the behaviors of mRNAs, proteins, protein interactions, biomodules, etc.). Integrate the different data types, graphically display them, and compare these to the model. Use hypothesis-driven approaches to explain the disparities between the model and experimental data. Driven by these hypotheses, generate a second round of perturbations and global analyses. Repeat this process iteratively, reformulating the model at each round, until theory

Table 6. Integrations of Systems Biology

| |
|---|
| biology/technology/computation |
| cross-disciplinary faculty |
| new technologies/high-throughput platforms |
| various levels of biological information |
| perturbations/modeling |
| graphical display/modeling/reengineering |
| discovery science/hypothesis-driven science |
| systems biology/education |
| academia/industry |

(model) and experimental data are in accordance. Thus, the essence of systems biology is this iterative and integrative approach.

• Develop a mathematic model that will accurately describe the behavior of the system—its systems or emergent properties—given any particular perturbation.

• Reengineer the system. Once the regulatory networks and biomodules (molecular machines) of a particular system are well understood, there will be the possibility of reengineering the system to enhance its systems properties (e.g., higher quality food, higher yield of food, etc.) or, indeed, to redesign it to carry out completely new functions.

We have successfully employed systems approaches for two biological systems: galactose utilization in yeast[55] and endomesodermal development in the sea urchin.[56] We have also written one of the first reviews on systems biology.[57]

The four major challenges facing biologists wishing to use systems approaches are as follows: (1) to develop more effective tools for genetic and environmental perturbations; (2) to access the powerful high-throughput tools necessary to capture information from the various hierarchical levels; (3) the ability to handle and integrate the large-scale data sets from the various hierarchical levels of biological information; and (4) the ability to formulate and optimize mathematical models of the systems employing the iteratively integrated information. These are all challenges that we are currently working on at the Institute for Systems Biology. It is likely that systems biology will be one of the major driving forces in biology of the 21st century.

**Implementation.** We started the Institute for Systems Biology to create, apply, and disseminate systems biology. We have 9 cross-disciplinary faculty (astrophysics, computer science, engineering, mathematics, biology, and chemistry) and have grown from a staff of 2 to 170 in two years. We have established genomic (DNA sequencing, genotyping, DNA arrays), proteomic, and cell-sorting high-throughput facilities, and we have a very strong computational infrastructure. We have a variety of industrial and academic partnerships. The most challenging aspects of systems biology involve the various integrations (Table 6). Perhaps the largest integration challenge is the effective interactions of biologists, chemists, computer scientists, engineers, mathematicians, physicians, and physicists. It is critical that non-biologists learn biology in a deep way, for their ability to contribute will be proportional to their level of understanding. The same is true for biologists—they must understand what engineering, mathematics, and physics can potentially contribute to biology. How does one integrate new technologies into preexisting high-throughput platforms? or acquire the space, resources, and talent to build these platforms? A striking challenge is the integration of different levels of biological information and its graphical display, modeling, and reengineering. Reengineering means rebuilding biological circuitry for designated objectives. Integrating discovery and hypothesis-driven science is the key to integrating small groups of investigators with the enormous opportunities of systems biology. Ultimately, the objective must be to integrate biology, medicine, computation, and technology. This will require challenging changes in how we educate scientists. Academia and industry will find these integrations the biggest challenge in the execution of systems biology.[58,59]

Systems biology and the global access to human variability (polymorphisms) lead to a fundamental revolution in medicine—predictive and preventive medicine.

## Predictive and Preventive Medicine Will Transform the Practice of Medicine

On average, 1 in 500 bp is polymorphic in comparing the genomes of two humans. Most of these polymorphisms do not affect the phenotype of the individual. However, a few encode differences in physiology (e.g., tall/short or thin/fat); a few others encode predispositions to disease. It is clear that the predispositions to most diseases are actually encoded by multiple genes.

My prediction is that over the next 10−15 years, certain polymorphisms in hundreds of genes will be correlated with predispositions for cancer, cardiovascular disease, immunologic disease, etc. At that time, physicians will be able to isolate DNA from the blood of the young, analyze the hundreds of potential disease-predisposing genes, and determine a probabilistic health history for each individual—this is **predictive** medicine. In 15−20 years, with the aid of systems biology, physicians will be able to understand the biological networks within which the defective genes reside and, hence, ascertain how to block the effects of these defective genes—this is **preventive** medicine. Since most individuals will have differing combinations of predisposing genes, medicine will become highly individualized. The transition to predictive/preventive medicine will occur over the next 15−20 years and will likely add 10 to 30 years to the productive lifespan of most individuals.

The Institute for Systems Biology is working on a variety of systems approaches to disease, including the multiparameter analysis of blood samples. In predictive and preventive medicine, diagnostics will be inextricably linked initially to therapy and eventually to prevention.

## Conclusion

I close with two points. First, if we return to Freeman Dyson's quotes (Figure 1), we see that our new technologies have opened up striking new opportunities—the DNA and protein sequencers and synthesizers and ink-jet oligonucleotide synthesizer have transformed biology and medicine. Indeed, they have helped catalyze two of the major paradigm changes in 21st century biology—systems biology and predictive and preventive medicine. Second, my career in technology development has necessarily expanded into areas concerning the organization of science—first in creating the novel cross-disciplinary Department of Molecular Biotechnology, and more recently the independent, nonprofit Institute for Systems Biology, whose mission is the development, application, and dissemination of systems approaches to biology and medicine. The integrations of systems biology are challenging, but the opportunities are limitless. Systems biology will continue to be driven by technological advances—in instrumentation, in strategies, and in integrating and modeling of biological information with yet to be invented computational tools.

## References

(1) Early, P.; Huang, H.; Davis, M.; Calame, K.; Hood, L. An Immunoglobulin Heavy Chain Variable Region Gene is Generated from Three Segments of DNA: $V_H$, D, and $J_H$. *Cell* **1980**, *19*, 981−992.
(2) Rogers, J.; Early, P.; Carter, C.; Calame, K.; Bond, M.; Hood, L.; Wall, R. Two mRNAs with Different 3′ Ends Encode Membrane-Bound and Secreted Forms of Immunoglobulin $\mu$ Chain. *Cell* **1980**, *20*, 303−312.
(3) Early, P.; Rogers, J.; Davis, M.; Calame, K.; Bond, M.; Wall, R.; Hood, L. Two mRNAs can be Produced from a Single Immunoglobulin $\mu$ Gene by Alternative RNA Processing Pathways. *Cell* **1980**, *20*, 313−319.
(4) Crews, S.; Griffin, J.; Huang, H.; Calame, K.; Hood, L. A Single VH Gene Segment Encodes the Immune Response to Phosphorylcholine: Somatic Mutation is Correlated with the Class of the Antibody. *Cell* **1981**, *25*, 59−66.
(5) Steinmetz, M.; Moore, K. W.; Frelinger, J. G.; Sher, B. T.; Shen, F.-W.; Boyse, E. A.; Hood, L. A Pseudogene Homologous to Mouse Transplantation Antigens: Transplantation Antigens are Encoded by Eight Exons That Correlate with Protein Domains. *Cell* **1981**, *25*, 683−692.
(6) Kim, S.; Davis, M.; Sinn, E.; Patten, P.; Hood, L. Antibody Diversity: Somatic Hypermutation of Rearranged VH Genes. *Cell* **1981**, *27*, 573−581.
(7) Steinmetz, M.; Winoto, A.; Minard, K.; Hood, L. Clusters of Genes Encoding Mouse Transplantation Antigens. *Cell* **1982**, *28*, 489−498.
(8) Crews, S.; Barth, R.; Hood, L.; Prehn, J.; Calame, K. Mouse c-myc Oncogene is Located on Chromosome 15 and Translocated to Chromosome 12 in Plasmacytomas. *Science* **1982**, *218*, 1319−1321.
(9) Steinmetz, M.; Minard, K.; Horvath, S.; McNicholas, J.; Frelinger, J.; Wake, C.; Long, E.; Mach B.; Hood, L. A Molecular Map of the Immune Response Region from the Major Histocompatibility Complex of the Mouse. *Nature* **1982**, *300*, 35−42.
(10) Siu, G.; Clark, S. P.; Yoshikai, Y.; Malissen, M.; Yanagi, Y.; Strauss, E.; Mak, T. W.; Hood, L. The Human T Cell Antigen Receptor is Encoded by Variable, Diversity, and Joining Gene Segments that Rearrange to Generate a Complete V Gene. *Cell* **1984**, *37*, 393−401.
(11) Koop, B. F.; Hood, L. Striking Sequence Similarity Over Almost 100 kilobases of Human and Mouse T-Cell Receptor DNA. *Nature Genet.* **1994**, *7*, 48−53.
(12) Rowen, L.; Koop, B. F.; Hood, L. The Complete 685 kb DNA Sequence of the Human $\beta$ T Cell Receptor Locus. *Science* **1996**, *272*, 1755−1762.
(13) Aguado, B.; Bahram, S.; Beck, S.; Campbell, R. D.; Forbes, S.; Geraghty, D.; Guillaudeux, T.; Hood, L.; Horton, R.; Inoko, H.; Janer, M.; Jasoni, C.; Madan, A.; Milne, S.; Neville, M.; Oka, A.; Qin, S.; Ribas-Despuig, G.; Rogers, J.; Rowen, L.; Shiina, T.; Spies, T.; Tamiya, G.; Tashiro, H.; Trowsdale, J.; Vu, Q.; Williams, L.; Yamazaki, M. Complete Sequence and Gene Map of a Human Major Histocompatibility Complex (MHC). *Nature* **1999**, *401*, 921−923.
(14) Glusman, G.; Rowen, L.; Lee, I.; Boysen, C.; Roach, J. C.; Smit, A. F.; Wang, A. K.; Koop, B. F.; Hood, L. Review: Comparative Genomics of the Human and Mouse T-Cell Receptor Loci. *Immunity* **2001**, *15*, 337−349.
(15) Hunkapiller, M. W.; Hood, L. New Protein Sequenator with Increased Sensitivity. *Science* **1980**, *207*, 523−525.
(16) Hewick, R. M.; Hunkapiller, M. W.; Hood, L. E.; Dreyer, W. J. A Gas−Liquid Solid-Phase Peptide and Protein Sequenator. *J. Biol. Chem.* **1981**, *256*, 7990−7997.
(17) Doolittle, R. F.; Hunkapiller, M. W.; Hood, L. E.; Devare, S. G.; Robbins, K. C.; Aaronson, S. A.; Antoniades, H. N. Simian Sarcoma *onc* gene, v-*sis*, is Derived from the Gene (or Genes) Encoding a Platelet-Derived Growth Factor. *Science* **1983**, *221*, 275−277.
(18) Prusiner, S. B.; Groth, D. F.; Bolton, D. C.; Kent, S. B. H.; Hood, L. E. Purification and Structural Studies of a Major Scrapie Prion Protein. *Cell* **1984**, *38*, 127−134.
(19) Oesch, B.; Westaway, D.; Wälchli, M.; McKinley, M. P.; Kent, S. B. H.; Aebersold, R.; Barry, R. A.; Tempst, P.; Teplow, D. B.; Hood, L. E.; Prusiner, S. B.; Weissmann, C. A. A Cellular Gene Encodes Scrapie PrP 27−30 Protein. *Cell* **1985**, *40*, 735−746.
(20) Hunkapiller, M. W.; Hood, L. E. Human Fibroblast Interferon: Amino Acid Analysis and Amino Terminal Amino Acid Sequence. *Science* **1980**, *207*, 525−526.
(21) Zoon, K. C.; Smith, M. E.; Bridgen, P. J.; Anfinsen, C. B.; Hunkapiller, M. W.; Hood, L. E. Amino Terminal Sequence of the Major Component of Human Lymphoblastoid Interferon. *Science* **1980**, *207*, 527−528.
(22) Raftery, M. A.; Hunkapiller, M. W.; Strader, C. D.; Hood, L. E. Acetylcholine Receptor: Complex of Homologous Subunits. *Science* **1980**, *208*, 1454−1457.
(23) Kent, S. B.; Hood, L. E.; Beilan, H.; Marriot, M.; Meister, S.; Geiser, T. A Novel Approach to Automated Peptide Synthesis Based on New Insights into Solid Phase Chemistry. Peptide Chemistry 1984. *Proceedings of the Japanese Peptide Symposium*; Isymiya, N., Ed.; Protein Research Foundation: Osaka, 1984; pp 217−222.
(24) Clark-Lewis, I.; Hood, L. E.; Kent, S. B. H. Role of Disulfide Bridges in Determining the Biological Activity of Interleukin 3. *Proc. Nat. Acad. Sci. U.S.A.* **1988**, *85*, 7897−7901.
(25) Párraga, G.; Horvath, S. J.; Eisen, A.; Taylor, W. E.; Hood, L.; Young, E. T.; Klevit, R. E. Zinc-Dependent Structure of a Single-Finger Domain of Yeast ADR1. *Science* **1988**, *241*, 1489−1492.
(26) Horvath, S. J.; Firca, J. R.; Hunkapiller, T.; Hunkapiller, M. W.; Hood, L. An Automated DNA Synthesizer Employing Deoxynucleoside 3′ Phosphoramidites. *Methods Enzymol.* **1987**, *154*, 314−326.
(27) Smith, L. M.; Sanders, J. Z.; Kaiser, R. J.; Hughes, P.; Dodd, C.; Connell, C. R.; Heiner, C.; Kent, S. B. H.; Hood, L. E. Fluorescence Detection in Automated DNA Sequence Analysis. *Nature* **1986**, *321*, 674−679.
(28) Blanchard, A. P.; Kaiser, R. J.; Hood, L. E. High-Density Oligonucleotide Arrays. *Biosens. Bioelectron.* **1996**, *11*, 687−690.
(29) Blanchard, A. P.; Hood, L. Sequence to Array: Probing the Genome's Secrets. *Nature BioTechnol.* **1996**, *14*, 1649.
(30) Strauss, E. C.; Kobori, J. A.; Siu, G.; Hood, L. E. Specific Primer-Directed DNA Sequencing. *Anal. Biochem.* **1986**, *154*, 353−360.
(31) Landegren, U.; Kaiser, R.; Sanders, J.; Hood, L. A Ligase-Mediated Gene Detection Technique. *Science* **1988**, *241*, 1077−1080.
(32) Boysen, C.; Simon, M. I.; Hood, L. Analysis of the 1.1-Mb Human $\alpha/\delta$ T-Cell Receptor Locus with Bacterial Artificial Chromosome Clones. *Genome Res.* **1997**, *7*, 330−338.
(33) Olson, M.; Hood, L.; Cantor, C.; Botstein, D. A Common Language for Physical Mapping of the Human Genome. *Science* **1989**, *245*, 1434−1435.
(34) Venter, J. C.; Smith, H.; Hood, L. A New Strategy for Genome Sequencing. *Nature* **1996**, *381*, 364.
(35) Mahairas, G. G.; Wallace, J. C.; Smith, K.; Swartzell, S.; Holzman, T.; Keller, A.; Shaker, R.; Furlong, J.; Young, J.; Zhao, S.; Adams, M. D.; Hood, L. Sequence-Tagged Connectors: A Sequence Approach to Mapping and Scanning the Human Genome. *Proc. Nat. Acad. Sci. U.S.A.* **1999**, *96*, 9739−9744.
(36) Roach, J. C.; Boysen, C.; Wang, K.; Hood, L. Pairwise End Sequencing: A Unified Approach to Genomic Mapping and Sequencing. *Genomics* **1995**, *26*, 345−353.
(37) Siegel, A. F.; Trask, B.; Roach, J. C.; Mahairas, G. G.; Hood, L.; van den Engh, G. Analysis of Sequence-Tagged-Connector Strategies for DNA Sequencing. *Genome Res.* **1999**, *9*, 297−307.
(38) Siegel, A. F.; van den Engh, G.; Hood, L.; Trask, B.; Roach, J. C. Modeling the Feasibility of Whole Genome Shotgun Sequencing Using a Pairwise End Strategy. *Genomics* **2000**, *68*, 237−246.
(39) Silver, J.; Hood, L. Automated Microsequence Analysis by Use of Radioactive Phenylisothiocyanate. *Anal. Biochem.* **1975**, *67*, 392−396.
(40) Hunkapiller, M. W.; Hood, L. E. Direct Microsequence Analysis of Polypeptides Using an Improved Sequenator, a Nonprotein Carrier (Polybrene), and High Pressure Liquid Chromatography. *Biochemistry* **1978**, *17*, 2124−2133.
(41) Hunkapiller, M. W.; Lujan, E.; Ostrander, F.; Hood, L. Isolation of Microgram Quantities of Proteins from Polyacrylamide Gels for Amino Acid Sequence Analysis. *Methods Enzymol.* **1983**, *91*, 227−236.
(42) Aebersold, R. H.; Teplow, D. B.; Hood, L. E.; Kent, S. B. H. Electroblotting onto Activated Glass: High Efficiency Preparation of Proteins from Analytical Sodium Dodecyl Sulfate−Polyacrylamide Gels for Direct Sequence Analysis. *J. Biol. Chem.* **1986**, *261*, 4229−4238.
(43) Aebersold, R.; Teplow, D. B.; Hood, L. E.; Kent, S. B. H. Electroblotting from Immobiline Isoelectric Focusing Gels for Direct Protein Sequence Determination. *Peptides Biol. Fluids* **1986**, *34*, 715−718.
(44) Kent, S.; Hood, L.; Aebersold, R.; Teplow, D.; Smith, L.; Farnsworth, V.; Cartier, P.; Hines, W.; Hughes, P.; Dodd, C. Approaches to Sub-Picomole Protein Sequencing. *BioTechniques* **1987**, *5*, 314−321.

(45) Aebersold, R.; Pipes, G. D.; Wettenhall, R. E. H.; Nika, H.; Hood, L. E. Covalent Attachment of Peptides for High Sensitivity Solid-Phase Sequence Analysis. *Anal. Biochem.* **1990**, *187*, 56–65.

(46) Wettenhall, R. E. H.; Aebersold, R. H.; Hood, L. E. Solid-Phase Sequencing of $^{32}$P-Labeled Phosphopeptides at Picomole and Subpicomole Levels. *Methods Enzymol.* **1991**, *201*, 186–199.

(47) Mononen, I.; Heisterkamp, N.; Kaartinen, V.; Williams, J. C.; Yates, J. R., III; Griffin, P. R.; Hood, L. E.; Groffen, J. Aspartylglycosaminuria in the Finnish Population: Identification of Two Point Mutations in the Heavy Chain of Glycoasparaginase. *Proc. Nat. Acad. Sci. U.S.A.* **1991**, *88*, 2941–2945.

(48) Kaartinen, V.; Williams, J.; Tomich, J.; Yates, J. R., III; Hood, L.; Mononen, I. Glycosaparaginase from Human Leukocytes. Inactivation and Covalent Modification with Diazo-Oxonorvaline. *J. Biol. Chem.* **1991**, *266*, 5860–5869.

(49) Griffin, P. R.; Coffman, J. A.; Hood, L. E.; Yates, J. R., III. Structural Analysis of Proteins by Capillary HPLC Electrospray Tandem Mass Spectrometry. *Int. J. Mass Spectrom. Ion Processes* **1991**, *111*, 131–149.

(50) Yates, J. R., III; Zhou, J.; Griffin, P. R.; Hood, L. E. Computer Aided Interpretation of Low Energy MS/MS Mass Spectra of Peptides. *Tech. Protein Chem. II* **1991**, *46*, 477–485.

(51) Hunkapiller, M.; Kent, S.; Caruthers, M.; Dreyer, W.; Firca, J.; Giffin, C.; Horvath, S.; Hunkapiller, T.; Tempst, P.; Hood, L. A Microchemical Facility for the Analysis and Synthesis of Genes and Proteins. *Nature* **1984**, *310*, 105–111.

(52) Moore, R. C.; Lee, I. Y.; Silverman, G. L.; Harrison, P. M.; Strome, R.; Heinrich, C.; Karunaratne, A.; Pasternak, S. H.; Azhar Chishti, M.; Liang, Y.; Mastrangelo, P.; Wang, K.; Smit, A. F. A.; Katamine, S.; Carlson, G. A.; Cohen, F. E.; Prusiner, S. B.; Melton, D. W.; Tremblay, P.; Hood, L. E.; Westaway, D. Ataxia in Prion Protein (PrP)-Deficient Mice is Associated with Upregulation of the Novel PrP–Like Protein Doppel. *J. Mol. Biol.* **1999**, *292*, 797–817.

(53) International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, Center for Genome Research, The Sanger Centre, Washington University Genome Sequencing Center, US DOE Joint Genome Institute, Baylor College of Medicine HumanGenome Sequencing Center, RIKEN Genomic Sciences Center, Genoscope and CNRS UMR-8030, GTC Sequencing Center, Department of Genome Analysis, Institute of Molecular Biotechnology, Beijing Genomics Institute/Human Genome Center, Multimegabase Sequencing Center, The Institute for Systems Biology: L. Hood, L. Rowen, A. Madan, and Shizen Qin, Stanford Genome Technology Center, Stanford Human Genome Center, University of Washington Genome Center, Department of Molecular Biology, Keio University School of Medicine, University of Texas Southwestern Medical Center at Dallas, University of Oklahoma's Advanced Center for Genome Technology, Max Planck Institute for Molecular Genetics, Cold Spring Harbor Laboratory, Lita Annenberg Hazen Genome Center, GBF–German Research Centre for Biotechnology, Genome Analysis Group, Scientific Management: National Human Genomes Research Institute, U.S. National Institutes of Health, Office of Science, U.S. Department of Energy, The Wellcome Trust. Initial Sequencing and Analysis of the Human Genome. *Nature* **2001**, *409*, 860–921.

(54) International Human Genome Sequencing Consortium. A Physical Map of the Human Genome. *Nature* **2001**, *409*, 934–941.

(55) Ideker, T.; Thorsson, V.; Ranish, J. A.; Christmas, R.; Buhler, J.; Eng, J. K.; Bumgarner, R.; Goodlett, D. R.; Aebersold, R.; Hood, L. Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network. *Science* **2001**, *292*, 929–933.

(56) Davidson, E. H.; Rast, J. P.; Oliveri, P.; Ransick, A.; Calestani, C.; Yuh, C.-H.; Minokawa, T.; Amore, G.; Hinman, V.; Arenas-Mena, C.; Otim, O.; Brown, C. T.; Livi, C.; Lee, P. Y.; Revilla, R.; Rust, A. G.; Pan, Z. J.; Schilstra, M. J.; Clarke, P. J. C.; Arnone, M. I.; Rowen, L.; Cameron, R. A.; McClay, D. R.; Hood, L.; Bolouri, H. A Genomic Regulatory Network for Development. *Science* **2002**, *295*, 1669–1678.

(57) Ideker, T.; Galitski, T.; Hood, L. A New Approach to Decoding Life: Systems Biology. *Annu. Rev. Genomics Hum. Genet.* **2001**, *2*, 343–372.

(58) Aebersold, R.; Hood, L. E.; Watts, J. D. Equipping Scientists for the New Biology. *Nature BioTechnol.* **2000**, *18*, 359.

(59) Aderem A.; Hood, L. Immunology in the Post-Genomic Era. *Nature Immunol.* **2001**, *2*, 1–3.

PR020299F