Proper Citation:
BIOLOGICAL AND BIOLOGY-RELATED INFORMATION AS A
COMMODITY AND THE RISE OF BIOINFORMATICS

“If you have $10 dollars and I have $10. You give me your $10 and I give you my $10, we each
still have $10. There seems to be a ‘conservation of money’. If you have an idea and I have an
idea, you give me your idea and I give you my idea. We each now have two ideas. Thus ideas
propagate and the number of ideas multiplies.”

Hwa A. Lim, Opening Remark, Bioinformatics & Genome Research, Inner Harbor Hotel, 1996.

1 The Dynamic High Tech World

“Ships in harbor are safe, but this is not what ships are built for”.¹ This saying applies extremely well to business. Entrepreneurs and venture capitalists go into business using their opportunity cost and capital in hopes of earning a fair return on their investment. But not all businesses are sure bets. In all businesses, there is an intrinsic element of risk. In most cases, the risks are calculated, weighed and then committed in a way to the owners’ advantage. The rules may be changed during the operations of the business, either by exploiting the “weak” or by stretching the legal limits. When the stake is high, the return-capital ratio is also expected to be correspondingly higher. Otherwise, it is normally just not worth wagering the risk.

In recent years, we have seen too often that a seemingly healthy company suddenly has to undergo an abrupt metamorphosis, either to stay afloat or to stay profitable in this cutthroat competitive marketplace. The most recent example is Apple Computer during the first half of the 1990s. Apple Computer has been an excellent company in coming out with new ideas, but by the time the ideas get to the market, the effect of products becomes diffused. Of late, with corporate reengineering and the introduction of a new line of “more brain, less brawn” line of products including iMacs, Power Macintosh G3, and PowerBook G3, Apple Computer has reemerged out of the woods and regain their market position.

This rapid change in market thrust is true not only in the computer world, but is also true for biotechnology, pharmaceutical and healthcare sectors. In order to compete and advance in this competitive marketplace, the chances of succeeding will be enhanced if a company stays focused in the areas it is good at and in the areas in which it can effectively compete.

¹ Peter Burwash, The Key to Great Leadership, (Torchlight, New York, 1995).
2 Types Of Focused Companies

A few examples of successful focused companies are in order at this point. There are at least three broad categories of these types of companies: operationally efficient, service-oriented and product-driven companies. Let us look at each in turn.

2.1 Operationally Efficient Companies

There are quite a number of operationally efficient companies, including Wal-Mart, McDonald, Price-Costco, etc.

A classic example of an operationally inefficient company is the failed boutique business of Aamco. In the mid-1970s, Aamco, a leading wholesaler of auto transmission parts, decided to diversify by opening up a chain of gift boutiques called Plum Tree. The plan called for the same system of centralized buying and centralized reselling to the franchise. By centralizing, sales of each item could be monitored and the system kept the inventory low while still maintaining the levels of stock needed to support the sales. The program went off to a great start but soon declined. The system automatically ordered larger quantities of most frequently used transmission parts in order to maintain the flow of stock. The same system also reordered more gift items customers purchased most frequently. But transmission parts and boutiques were not the same: mechanics would usually repeat the same sort of repair day after day; on the other hand, customers would rarely buy the same gifts for their friends time after time. Presented with what they already had, customers stopped coming to the boutique, and the boutique business failed miserably.

In contrast, the success of Wal-Mart is transparent. When Sam Walton opened the first Wal-Mart store in 1962, it was the beginning of an American success story that no one could have predicted. A small-town merchant who had operated variety stores in Arkansas and Missouri, Walton was convinced that consumers would flock to a discount store with a wide array of merchandise and friendly service. At its core, Wal-Mart is a place where prices are low and customer service is high. Because Wal-Mart carefully controls expenses to maintain its low price structure, customers do not have to wait for a sale to realize savings. Backing up the hometown flavor of a Wal-Mart store is the industry's most efficient and sophisticated distribution system. The system allows each store to personalize the merchandise assortment to match the community’s needs.

---

2.2 Service-oriented Companies

The service sector now provides almost four-fifths of all jobs in the U.S., making it our largest job provider, our major producer of wealth. According to a study conducted by The Forum Corporation, a global leader in developing workplace learning solutions and in helping companies build lasting brand loyalty, 70% of those customers who change companies do so because of poor service. As we move into the new millennium, it is interesting to see how corporations become increasingly aware of the fact that it is a lot cheaper to keep a customer than to spend more money on advertising to gain a new one.

There are countless successful service-oriented companies, including Nordstrom, Home Depot, and Federal Express. The “Pony Express” will provide a good case study. In the mid twelfth century, the Pony Express of the Khan Dynasty could cover in just a day and a night what would normally have been a ten-day journey. Using the most powerful horses, the riders, wearing wide belts with many bells that announced their approach to the horse post houses. At twenty-five intervals along the route, a fresh horse would be saddled and waiting for the rider, who would press on, towards the Palace of Genghis Khan at Chard.

In more recent times, the short-lived Pony Express in the Wild West United States amounted to little more than a romantic footnote to the history of American frontier. During the brief period April 1860 and October 1861, a small band of brave and resourceful men provided a unique high-speed service. By changing horses every seventy-five miles, riding day and night, rain or shine, and evading hostile highway bandits, historical record shows that they could cover 2,000 miles between St. Joseph (Missouri) and Sacramento (California) in less than ten days. This was very impressive in those days. But they were soon driven out of business by a single strand of copper telegraph wire.

Today, Frederick Wallace Smith, an inventive entrepreneur, created a similar, very successful, counterpart - the Federal Express Company. The company provides fast delivery of documents while Western Union is cutting back. This example shows that speed is not the only factor. The telegraph is a technology that is alphanumeric, i.e., it works only with numbers, letters, and other keyboard characters. Federal Express, on the other hand, can deliver anything at all within legal limits. While efficient use of technology appears to have carried Federal Express to a clear lead over the telegraph, winners cannot rest on their laurels in any high technology world. A new machine, a desktop copier that can send graphical replicas of letter-sized documents, can now offer telegraphic speed document transmission. And we cannot ignore the effect of the Internet and the World Wide Web (WWW) on the industry, and the new regulation recognizing faxed signatures.

This is a good example of why in all areas of business, changes brought about by technology will leave some of today’s leader scrambling to catch up to more nimble competitors. Success usually comes to those who apply technology to their advantage.
2.3 Product-driven Companies

The United States is still by far the leading nation in computer technology and biotechnology. Most of these companies have their respective core products. Examples of these computer companies include Microsoft, Intel, Apple, etc. A few examples of product-driven biotechnology companies and their known product lines are: Affymetrix (DNA chip), Advanced Cell Technology (stem cells), Geron (stem cells and cell lines), Chiron (various products), Genentech (several products), Merck (many products). A noteworthy fact from the list is that as we go down the list, the number of products associated with the company increases, i.e., the company is more diversified.

In 1971, Intel introduced the world’s first microprocessor, which sparked a personal computer (PC) revolution that would change the world. About 75% of PC in use around the world today are based on Intel-architecture microprocessor. Today, Intel supplies the PC industry with computer chips, boards, systems, software that are components of the most popular computer architecture. They help create advanced and high performance computing systems for PC users. The product line includes: 1) microprocessor chips, especially the ubiquitous Pentium chips; 2) networking and communication products; 3) semiconductor products such as reprogrammable memory for cellular phones, heavy machinery parts. They supply not only to the computer industry, but they also hold market shares in: 1) manufacturers of computers and peripherals; 2) other manufacturers like makers of automobiles and a wide range of industrial and telecommunication equipment; 3) PC users who buy Intel enhancements for business communication, networking, etc; 4) scientists and engineers working on sophisticated computational problems. Through its Research Council, it champions university research to drive innovations and the advancement of technology in the computer industry so that there is a continuous stream of new product line. The timely response and the professionalism with which Intel handled the 6th decimal rounding error of Pentium chips in 1995 is yet another reason why the company is so successful.

3 Biology And Biology-Related Disciplines - Place In Science

The prevailing view is that biology has jostled to the center stage at the expense of the physical sciences. This is a fallacy.

As we close the twentieth century and begin the new millennium, if we look back on the twentieth century, we can conclude that its first half was shaped by the physical sciences but its second by biology. The first half brought the revolution in transportation, communication, mass production technology and the beginning of the computer age. It also, pleasantly or unpleasantly enough, brought in the nuclear

---

weapons and the irreversible change in the nature of warfare and environment. All of these changes and many more rested on physics and chemistry. Biology was also stirring over those decades. The development of vaccines and antibiotics, early harbingers of the green revolution are all proud achievements. Yet the public’s preoccupation with the physical sciences and technologies, and the immense upheavals in the human condition which these brought, meant that biology and medicine could only move to the center stage somewhat later. Moreover, the intricacies of living structures are such that their deepest secrets could only be revealed after the physical sciences had produced the tools - electron microscopes, radioisotopes, chemical analyzers, laser, DNA sequencers, and rather importantly, the computer - required for probing studies. Accordingly, it is only now that the fruits of biology have jostled their way to the front pages.

Computer technology, especially computational power, networking and storage capacity, has advanced to a stage that it is capable of handling some of the current challenges posed by biology. This makes it possible to handle the vast amount of biological and biology-related data that have been and are being generated as a result of genome projects, the dissemination of the data on the Internet, and provide the teraflop compute power required for complicated analyses to penetrate the deepest secrets of biology and biology-related disciplines. Consequently, the time was prime for a marriage made in heaven between biology and biology-related disciplines, and computer science in the late 1980s. Thus the birth of the eclectic bioinformatics in the 1990s.

Bioinformatics will continue to be ubiquitous and play a critical role way into the twenty-first century as biotechnology (BT), information technology (IT) and nanotechnology (NT) are converging to form a triumvirate of BIN convergence.

4 Pharmaceutical Companies As Product-Driven Companies

It is not coincidental that the ten largest pharmaceutical colossi are those which are most visionary: Glaxo Wellcome (UK, sales: $11.68M; capitalization: $45.17M), Merck & Co., (US, sales: $10.96M; capitalization: $83.92M), Novartis (Switzerland, sales: $10.94M; capitalization: $78.87M), Hoechst (Germany, sales: $9.42M; capitalization: $18.64M), Roche Holding (Switzerland, sales: $7.82M; capitalization: $67.21M), Bristol-Myers Squibb (US, sales: $7.81M; capitalization: $44.67M), Pfizer (US, sales: $7.07M; capitalization: $43.25M), SmithKline Beecham (UK, sales: $6.6M; capitalization: $29.61M), Johnson & Johnson (US, sales: $6.3M; capitalization: $63.17M), Pharmacia & Upjohn (UK, sales: $6.26M; capitalization: $22.14M). We intentionally post this slightly outdated 1998 chart for comparison with a more recent data. It should be noted how swiftly these companies jostle for relative positions. For example, AstraZeneca, which was

---

number two in the Y2K listing, comes in first in Y2001, having increased its market capitalization from $57 billion to $85 billion. Amgen slipped from number one last year ($75 billion) to second place ($68 billion). Big pharmaceutical companies grew appreciably as well. Pfizer, which held second place in the Y2K listing with a market capitalization of $125 billion, tops the Y2001 pharmaceutical company chart at $264 billion. Merck & Co. fell from first in Y2K ($140 billion) to third ($166 billion) in Y2001.

Table 1. Top 10 pharmaceutical companies. (Source, Telescan Prosearch, No. 5, May 25, 2001).

<table>
<thead>
<tr>
<th>Company</th>
<th>Market Cap (Billion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pfizer Inc</td>
<td>264.3</td>
</tr>
<tr>
<td>GlaxoSmithKline plc</td>
<td>167.4</td>
</tr>
<tr>
<td>Merck &amp; Co. Inc</td>
<td>166.5</td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>135.9</td>
</tr>
<tr>
<td>Bristol-Myers Squibb Co.</td>
<td>105.3</td>
</tr>
<tr>
<td>Eli Lilly &amp; Co.</td>
<td>94.4</td>
</tr>
<tr>
<td>American Home Products Corp.</td>
<td>80.2</td>
</tr>
<tr>
<td>Abbott Laboratories</td>
<td>79.6</td>
</tr>
<tr>
<td>Pharmacia Corp.</td>
<td>63.2</td>
</tr>
<tr>
<td>Schering-Plough Corp.</td>
<td>60.9</td>
</tr>
</tbody>
</table>

Table 2. Top 50 biotechnology companies. (Source: Telescan Prosearch, No. 5, May 25, 2001)

<table>
<thead>
<tr>
<th>Position</th>
<th>Company</th>
<th>Position</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AstraZeneca plc</td>
<td>26</td>
<td>OSI Pharmaceuticals Inc.</td>
</tr>
<tr>
<td>2</td>
<td>Amgen Inc.</td>
<td>27</td>
<td>Airon</td>
</tr>
<tr>
<td>3</td>
<td>Genentech Inc.</td>
<td>28</td>
<td>Myriad Genetics Inc.</td>
</tr>
<tr>
<td>4</td>
<td>Serono SA</td>
<td>29</td>
<td>Techne Corp</td>
</tr>
<tr>
<td>5</td>
<td>Genzyme Corp.</td>
<td>30</td>
<td>QLT Inc.</td>
</tr>
<tr>
<td>6</td>
<td>Chiron Corp.</td>
<td>31</td>
<td>Tanox Inc.</td>
</tr>
<tr>
<td>7</td>
<td>Celltech Group plc</td>
<td>32</td>
<td>Regeneron Pharmaceuticals</td>
</tr>
<tr>
<td>8</td>
<td>Biogen Inc.</td>
<td>33</td>
<td>Cell Therapeutics Inc.</td>
</tr>
<tr>
<td>9</td>
<td>IDEC Pharmaceuticals Corp.</td>
<td>34</td>
<td>Albany Molecular Research Inc.</td>
</tr>
<tr>
<td>10</td>
<td>Immunex Corp.</td>
<td>35</td>
<td>Titan Pharmaceuticals Inc.</td>
</tr>
<tr>
<td>11</td>
<td>MedImmune, Inc.</td>
<td>36</td>
<td>CV Therapeutics Inc.</td>
</tr>
<tr>
<td>12</td>
<td>Human Genome Sciences, Inc.</td>
<td>37</td>
<td>Scios Inc.</td>
</tr>
<tr>
<td>13</td>
<td>Applera Corp.</td>
<td>38</td>
<td>Genencor International Inc.</td>
</tr>
<tr>
<td>14</td>
<td>Gilead Sciences Inc.</td>
<td>39</td>
<td>Xoma Ltd.</td>
</tr>
<tr>
<td>15</td>
<td>Abgenix Inc.</td>
<td>40</td>
<td>Trimeris Inc.</td>
</tr>
<tr>
<td>16</td>
<td>ImClone Systems Inc</td>
<td>41</td>
<td>Cubist Pharmaceuticals Inc.</td>
</tr>
<tr>
<td>17</td>
<td>Icos Corp.</td>
<td>42</td>
<td>Immunomedics Inc.</td>
</tr>
<tr>
<td>18</td>
<td>Cephalon Inc.</td>
<td>43</td>
<td>Enzo Biochem Inc.</td>
</tr>
<tr>
<td>19</td>
<td>Enzon Inc.</td>
<td>44</td>
<td>Ligand Pharmaceuticals Inc.</td>
</tr>
<tr>
<td>20</td>
<td>Applera Corp.</td>
<td>45</td>
<td>Exelixis Inc.</td>
</tr>
<tr>
<td>21</td>
<td>Sepracor Inc.</td>
<td>46</td>
<td>Corixa Corp</td>
</tr>
<tr>
<td>22</td>
<td>Vertex Pharmaceuticals Inc.</td>
<td>47</td>
<td>Amylin Pharmaceuticals Inc.</td>
</tr>
<tr>
<td>23</td>
<td>Celgene Corp.</td>
<td>48</td>
<td>Arena Pharmaceuticals Inc.</td>
</tr>
<tr>
<td>24</td>
<td>Medarex Inc.</td>
<td>49</td>
<td>Biotechnology General Corp.</td>
</tr>
</tbody>
</table>
Each of these companies has a well-defined mission and each has various core products. For example, the theme of Bristol-Myers-Squibb is to be in business whose products help to enhance and extend human life. Pfizer has drug of the decade – and Merck has Propecia to treat hair loss. Similarly, Bayer is a diversified, international chemical and pharmaceutical company. Among the best-known names are undoubtedly the drug of the century - Aspirin, and Alka-Seltzer. A good 50% of the company’s sales are attributed to products developed in its own research laboratories in the last 15 years.

In the aftermath of the World Trade Center Twin Tower incident of September 11, 2001, we anticipate that companies, such as Eli Lilly, which produce medicines for anthrax and other bio agents to move up the rank.

5 The Economics Of Drug Discovery

Let us digress a little and look at drug discovery as a case example before we return to bioinformatics. Of about 5,000-10,000 compounds studied, only one drug gets onto the market. Even if the compounds make it to clinical trials, only 1 in 5 gets to the market. In the discovery phase, each drug costs about $156 million. The FDA processes I, II & III cost another $75 million. This brings the total to about $231 million (1994 figure. The 2000 figure is about $400 million) for each drug put onto the market for consumers. The time required for approval is equally long, as shown in the figure. These phases constitute parts of the manufacturing, regulatory and cost factors of drug discovery.

By the time a drug gets on to the shelf of a pharmacy, it will have been 15 years since preclinical testing, and about $400-$700 million will have been spent.

---

6 Future Pharmaceutical Discoveries

Traditionally, large pharmaceutical companies have a cautious, mostly chemistry- and pharmacology-based approach to the discovery and preclinical development program and therefore, do not yet have expertise in-house to generate, evaluate, and manage genetic data. Accordingly, many purchase licenses to tap into commercially generated and managed genomic databases. Until large companies build up their expertise and experience in the direction of genomics and functional genomics, there is a target of opportunity for smaller, more adept and savvy companies to fill the gap. Undoubtedly, when larger companies become more sophisticated and comfortable with current technology, the next generation technology-based opportunities will be in the hands of the nimble, more flexible smaller companies to again cycle and leverage for revenues from their larger counterparts. To constantly stay one step ahead is a winning and survival game plan for small companies.

The accepted consensus is that future pharmaceutical discoveries will stem from biological and biology-related information. Major pharmaceutical companies develop new core products. These companies are either slower in response, do not want to develop full scale sequencing expertise and maintain huge proprietary database in-house (except those pertinent to their product lines), or they do not want to commit the financial resources for such purposes.

But to stay competitive, these companies have to respond quickly and do need access to comprehensive genetic, biological and biology-related information for timely and accurate decision making. Thus, since 1996, we have seen bioinformatics divisions springing up in all major pharmaceutical companies to either partake in this exciting new area, or to partner with smaller, more nimble research institutions or companies. Because of this, smaller companies are
constantly being formed to take advantage of the window of opportunities, some of which survive, and many flounder. In general, these small companies try to develop technologies – platform hardware (for example, AbMetrix, Affymetrix and Axon) and software (for example, InforMax and GeneGo), produce a database of some form and then generate revenue from the database by either selling subscriptions to the database, or selling information derived from the database (for example, Celera, deCode and Incyte Genomics).

The current practice to cut down on research and development (R&D) costs favors outsourcing or farming out certain aspects of R&D to smaller companies, universities, and even abroad, for example China and India, where the talent pool is excellent and the labor cost is still very attractive.

Other options include forming strategic alliances and forming consortia such as the Single Nucleotide Polymorphism Consortium.

7 The Rise Of Bioinformatics

On May 3rd, 1996, a BioMASS panel, in conjunction with the BioSCIENCE Career Fair and sponsored by American Association for the Advancement of Science (AAAS), was held at Stanford University Medical Center. The title of the panel discussion was “Ph.D. Promises: The future of the Bioscientist in Academia and Industry”.

There were six panelists: P. Gardner, M.D. (VP Research, Alza Corporation & Associate Professor at Stanford University); R. Grymes, Ph.D. (Manager, Outreach Program, NASA); H.A. Lim (HAL), Ph.D. (then Director, Bioinformatics, Hyseq); J. MacQuitty, Ph.D., M.B.A. (CEO, GenPharm International); T. Schall, Ph.D., (Senior Scientist, DNAX Research Institute); J. Shaw, Ph.D. (Founder, The Stable Network); and R. Simoni, Ph.D. (Professor at Stanford and panel discussion moderator). HAL was the bioinformatics panelist. He was rather taken unexpectedly that so many Ph.D. students, post-doctoral fellows and young audience asked him what bioinformatics was. There were even individuals who commented that the bioinformatics community must be extremely small, so small that they had not heard of it until the panel discussion.

In the following few weeks, the whole matter took a drastic swing, mainly due to AAAS and the Science magazine. Science magazine published a series of articles on the subject and this brought the area a lot of attention and publicity.

Bioinformatics is a rather young discipline, bridging the life and computer sciences. The need for this interdisciplinary approach to handle biological knowledge is not insignificant. It underscores the radical changes in quantitative as well as qualitative terms that the biosciences have been seeing in the last two decades or so. The need implies:

---

* The meeting was announced in Science, May Issue, 1996.
our knowledge of biology and biology-related disciplines has exploded in such a way that we need powerful tools to organize the knowledge itself;

the questions we are asking of biological and biologically-related systems and processes today are getting more sophisticated and complex so that we cannot hope to find answers within the confines of unaided human brains alone.

The current functional definition of bioinformatics is “the study of information content and information flow in biological and biology-related systems and processes.” It has evolved to serve as a bridge between the observations (data) in diverse biologically-related disciplines, the derivations of the understanding (information) about how the systems or processes function, and subsequently the application (knowledge). In more pragmatic terms of a disease, bioinformatics is the creation of a registry, unraveling of the dysfunction, and the subsequent search for a therapy.

7.1 The Beginning

The interest in using computers to solve challenging biological problems started in the 1970s, primarily at Los Alamos National Laboratory, and pioneered by Charles DeLisi and George Bell. Among the team of scientists were Michael Waterman, Temple Smith, Minoru Kanehisa, Walter Goad, Paul Stein and Gian Carlo Rota.

In the late 1980s, following the pioneering work of DeLisi and Bell, Hwa A. Lim (HAL) realized the significance of marrying computer science and biology, he tried to come up with a captivating word for the area. In 1988, HAL first coined the word “bio-informatique”, with a little bit of French flavor, to denote the subject area. A year later, after making a few surveys, the preparation process for convening the first international conference was begun to bring awareness to the community. The process started by electronic mail since interested participants were most likely computer-savvy. During those days, the smtp mailers on VAX machines (the good hey days of VMS operating system) were not very forgiving. Email frequently bounced because of the hyphen in “bio-informatique”. In an attempt to overcome the problem, the word was changed to “bio/informatique”, allowing for the fact that the “bio” was just a prefix that could be substituted. The change did not help. Similar email problems persisted, which prompted HAL to drop the “/” completely and the word assumed the form “Bioinformatique”. Two conference secretariats commented that the word was a little too French. It was then appropriately modified to “Bioinformatics: Coining of”, in conformance with subjects like “optics”, “statistics”, “mathematics”...

Photo 1. A group photo of a subset of the 120 participants of the First Bioinformatics Conference. Notables: Alexander Bayev (Chairman, USSR Human Genome Project), George Bell (Acting Director, Los Alamos National Laboratory Genome Project), Charles Cantor (US Department of Energy Human Genome Program Principal Scientist), Anthony Carrano (Director, Lawrence Livermore Genome Project), Charles DeLisi (with Senator Domenici of New Mexico, an early proponent of the Greatest Wellness Project, which would be later called the Human Genome Project), Michel Durand (French Attache), a 6-member delegate from Japan (including RIKEN), and a huge 14-member Soviet delegate, something very unusual during the then Cold War era. (April 1990, Tallahassee, Florida).

The very first ever international conference on bioinformatics was chaired and organized by the author (HAL), with help from Professor Charles R. Cantor, then Chairman of the College of Physicians & Surgeons at Columbia University; and late Professor Joseph E. Lannutti, then Director of Supercomputer Computations Research Institute at Florida State University. The first conference was held at the Florida State Conference Center, Tallahassee, from April 10 through 13, 1990. Notable among the participants were: Charles DeLisi (Dean, College of Engineering, Boston University), Charles Cantor (then Director, Lawrence Berkeley National Laboratory Genome Program), George Bell (then Acting Director, Los Alamos National Laboratory Genome Program), Anthony Carrano (Director, Lawrence Livermore National Laboratory Genome Program), Temple Smith (then Director at Dana Farber Cancer Center of Harvard Medical School), late Alexander Bayev (then Chairman, USSR Genome Program), Boris Kaloshin (USSR Dept. of Sc. & Tech), M. Durand (French Attaché), N. Shimizu (Head, Department of Molecular Biology, Keio University School of Medicine), I. Endo (RIKEN, Japan), N. Nordén (Sweden), and others (120 participants in total). The conference was funded by U.S. Department of Energy, and Florida Technology Research and Development Authority, Thinking Machines Corp., Digital Equipment Corp., CRAY
Research Inc. A proceeding volume was compiled.\textsuperscript{11} Note that the sponsors were primarily federal and state agencies, and general-purpose computer companies. Note also the huge 14-member USSR delegate headed by Academician Alexander Bayev, a rather unusual phenomenon during the Cold War era. Also present was a 6-member delegate from RIKEN of Japan.

\textbf{Photo 2.} Hal, Dr. Charles DeLisi, Professor Charles Cantor, Academician Bayev, and Professor Joseph E. Lannutti at a press conference, Florida Press Club. Hal, who coined the word "bioinformatics" in 1987, was the Chairperson of this very first international conference on bioinformatics. Dr. Charles DeLisi, one of the early proponents of the Human Genome Project, was Dean of College of Engineering, Boston University. Professor Charles Cantor was US DoE Human Genome Project Principal Scientist (present: CSO, Sequanom, San Diego). Late Academician Bayev (1904-1994) was Chairman of the USSR Human Genome Project. Late Professor Lannutti (1926-1998) was initiator of the Supercomputer Computations Research Institute, host of the conference. (April 1990, Tallahassee, Florida).

7.2 \textit{Subsequent Years}

The conference series continued and The Second International Conference on Bioinformatics, Supercomputing and Complex Genome Analysis took place at the TradeWinds Hotel, St. Petersburg Beach, Florida, from June 4 through 7, 1992. This conference was originally planned for St. Petersburg (Leningrad), USSR. The breakup of the Former Soviet Union forced HAL to come up with an alternative plan in less than seven months. St. Petersburg Beach was chosen partly because of the

location, and partly because of its name (just like St. Petersburg of Russia). Participants from more than thirteen countries worldwide took part. A joke that circulated during and after the conference was that some attendees of the conference mistakenly went to St. Petersburg of Russia. The conference was partially funded by Intel Corp., MasPar Computer Corp., World Scientific Publishing Co., Silicon Graphics Corp., Florida Technological Research & Development Authority, U.S. Department of Energy, U.S. National Science Foundation. A second proceeding volume was edited to bring the subject area to the then relatively small community.  

Notable among the participants was a 12-member delegate from Genethon, France. Note the participation of federal and state agencies, special-purpose computer companies and publishing houses.

Photo 3. A group photo of a subset of the 150 participants of the Second Bioinformatics Conference. Notables: Christian Burks (Los Alamos National Lab., present: VP, Exelixis), Charles Cantor (Boston University, present: CSO, Sequenom Inc.), Rade Drmanac (Argonne National Lab, present: CSO, Hyseq, Inc.), Chris Fields (NCGR, present: VP, Applied Biosystems), Pavel Pevzner (University of Southern California, present: Ronald R. Taylor Chair Professor of Computer Science, University of California at San Diego), Temple Smith (Smith-Waterman Algorithm, Harvard University, present: Professor, Boston University), Robert Robbins (Johns Hopkins University, present: VP, Fred Hutchinson Cancer Center), Chris Sander (EMBL, present: Chief Information Technologist, Millennium), David Searles (University of Pennsylvania, present: VP, SmithKline Beecham), Mark Adams, Phil Green (known for his Phred and Phrap, inducted to the U.S. National Academy of Science, 2001), Edward Uberbacher (Grails), a delegate from Korea Institute of Science and Technology, a delegate from Japan, a delegate from Russia, and a huge 12-member delegate from France (including Genethon). (June 1992, St. Petersburg Beach, Florida).

The third conference, The Third International Conference on Bioinformatics & Genome Research, took place at Augustus Turnbull III Florida State Conference Center, Tallahassee, Florida, from June 1 through 4, 1994. It was partially funded by Compugen Ltd., Eli Lilly and Company, MasPar Computer Corp., World Scientific Publishing Co., Pergamon Press, U.S. Department of Energy, U.S. National Science Foundation, U.S. National Institutes of Health, International Science Foundation. The proceedings were gathered in a volume. A noteworthy point is that the sponsors were federal, state and international agencies, special-purpose computer companies, pharmaceutical companies and publishing houses.

Photo 4. A group photo of a subset of the 130 participants of the Third Bioinformatics Conference. Notable at the meeting was a panel discussion on sequence databases (Philip Bucher, James Fickett, Murray Smigel, Andrzej Konopka, Douglas Smith, Philippe Rigault, and Tom Slezak), and another panel discussion on technology transfer (Mike Devine, Assoc. VP of Research, Florida State University, Richard Hogg, Provost & VP of Academic Affairs, Florida A&M University, James Ludwig, Eli Lilly & Company, Richard MacDonald, Biosym Technologies). (June 1994, Tallahassee, Florida).

7.3 Bioinformatics Conference Going Commercial

Soon after the Third Conference, federal and state agencies went through a period of downsizing and streamlining. Cambridge Healthtech Institute (CHI) and the HAL decided that the conference series should go commercial and be self-supporting. Initial negotiations for CHI to take over the biennial conference series soon after the Conference in 1994. Judging from prior successes of the conference and the rising popularity of the subject area, CHI decided to make the conference series an annual event.

A noteworthy point is that even though the number of participants had been intentionally limited to less than 150 in the first three conferences, the number climbed steadily to 350 in the Fifth Conference, a clear indicator and good measure of the increasing popularity of the subject area. The Opening Ceremony of the Tenth Anniversary of the Conference Series, held at the Fairmont Hotel, San Francisco, on Father's Day Sunday evening (June 17, 2001) and chaired by HAL, was attended by 700 participants in 2001. Through the conference, the number of attendance remained the same, and there were 50 corporate sponsors and exhibitors.
8 Genomic Companies As Service-Oriented Companies

Many genomics companies have unique, high-throughput, cost effective technology to do sequencing and to collect biological or biology-related data. But, as shown in the table, data is not commercializable, but information is. This leads naturally to a conceptual flowchart of biological or biology-related data, as depicted in the figure. Or in terms of physical design, the corresponding databases as illustrated.

Table 3. A table to compare and contrast data and information.

<table>
<thead>
<tr>
<th>Data Are</th>
<th>Information is</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stored fact</td>
<td>Presented fact</td>
</tr>
<tr>
<td>Inactive (they exist)</td>
<td>Active (enables doing)</td>
</tr>
<tr>
<td>Technology-based</td>
<td>Business-based</td>
</tr>
<tr>
<td>Gathered from various sources</td>
<td>Transformed from data</td>
</tr>
</tbody>
</table>

Biodata

\[ \downarrow \]

Bioinformation

<table>
<thead>
<tr>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infobase</td>
</tr>
<tr>
<td>Knowledgebase</td>
</tr>
</tbody>
</table>

\[ \downarrow \]

Next generation genomics

Figure 2. A flowchart to show the paradigm of biodata. The prefix “bio” can very well be substituted for “chem”, “health” or any biology-related disciplines.

Figure 3. The paradigm of biodata presented in a more physical form, i.e., as various databases.

In order to maintain such a scheme, a possible strategic plan is outlined in Figure 4.14

---

Figure 4. A chart showing the flow and planning of information, in particular, bioinformation. The sequence is: assessment, strategy and execution.
8.1 **Bioinformatics - Mission and Goals**

Typically, some of the goals and missions of a bioinformatics division might include, among many other possibilities and combinations:
- To perform decision making by centering around intelligent interpretation of existing genetic information, for example, microarray profiles;
- To identify what information may yet be needed, define what may yet be done, for example, microarray data;
- To enable corporate partners to accelerate identification of genetic information for gene-based, small molecule-based, pathway-based drug targets;
- To validate this selection through sequencing-derived drug-genome interaction studies, for example, proteomics;
- To package this information for efficient decision making throughout a partner’s product development cycle.

The goals and mission may vary in accordance with needs, and very much driven by applications and clients.

![Diagram of drug discovery process](image)

**Figure 5.** The drug discovery process involves six major steps: target identification, screen development, screening, hit characterization, target validation, and lead optimization. Other supporting technologies can help in accelerating each of the processes. (Figure adapted from Dr. Don Halbert, Abbot Laboratories).
Figure 6. The time involved in each step of the drug discovery process is a very instructive piece of information. From a financial perspective, it is the effort that is involved is more indicative of the risk and cost involved. For example, lead optimization is a big bottleneck in drug discovery. (Figure adapted from Dr. Don Halbert, Abbot Laboratories).

A very seldom cited, yet very important, advantage of using informatics is that information is disciplinary-blind, that is, information can link disciplines that are seemingly unrelated to researchers. For example, it may seem that the pharmaceutical sector and the agriscience sector are not related. But if we compare and contrast drugs and seeds, certain similarities between the two sectors begin to reveal themselves.

Table 4. A comparison and contrast of the pharmaceutical and agriscience sectors.

<table>
<thead>
<tr>
<th>Pharma</th>
<th>Farmer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drugs are:</td>
<td>Seeds are:</td>
</tr>
<tr>
<td>- Chemicals</td>
<td>- Organisms</td>
</tr>
<tr>
<td>- Raw materials for screening</td>
<td>- Raw materials for breeding</td>
</tr>
<tr>
<td>- Sorted by chemistry</td>
<td>- Sorted by testing</td>
</tr>
<tr>
<td>New drug are:</td>
<td>New seeds:</td>
</tr>
<tr>
<td>- High margin</td>
<td>- Low margin</td>
</tr>
<tr>
<td>- Low volume</td>
<td>- High volume</td>
</tr>
<tr>
<td>Genes are:</td>
<td>Genes</td>
</tr>
<tr>
<td>- Linked to diseases</td>
<td>- Increase yield</td>
</tr>
</tbody>
</table>
Figure 7. How combinatorial chemistry informatics and genomics high throughput sequencing informatics have shifted the discovery paradigm in the pharmaceutical industry. (Figure adapted from Benjamin Bowen, Pioneer Hi-Bred International).
8.2 Bioinfobahn

Since bioinformatics is a marriage of computer and biology and biology-related disciplines, it is not surprising that it is well kept abreast with advances in computer technology, in particular, the Internet technology.

The ubiquitous computer network sprang from the mind of a research psychologist – J.C.R. Licklider - working for the Pentagon in the early 1960s. He envisioned what he called an Intergalactic Computer Network.\(^{15}\) The Pentagon was interested in part because it needed a communications system able to withstand nuclear attacks. A decentralized computer network would satisfy that requirement. Scientists were also intrigued by the idea of making telecommunications systems more efficient by breaking apart messages into “packets” of information that could be routed through different paths around the nation and reassembled by computers at the destination.

Although the technology had obvious commercial applications, the two telecommunications leaders at the time – IBM and AT&T – were not interested in participating because computers were not cheap enough to make such a network affordable. Instead, Pentagon’s Advanced Research Projects Agency and the civilian National Science Foundation nurtured the Internet for more than twenty-five years until 1994, when sufficient number of companies saw the potential of the Internet.

So, the Internet came into being about twenty-five years ago as a successor to ARPANET, a U.S. military network disguised to provide networking capabilities with a high redundancy. The principle behind has remained unchanged and has proven very powerful: to have every computer potentially talk to each other, regardless of what platform, what network path the communication actually takes.

By going cybernized, bioinformation and bioknowledge disseminate at a much timely rate. There are countless electronic publications on the Internet. Genome projects have produced a massive amount of data, resulting in over 400 individual databases at companies and institutions around the world. These publications appear in all formats - regular ASCII text, postscript, hypertext, Java and other derivations therefrom – supported by the Internet and the World Wide Web.

It is also fair to say that the Human Genome Project would not have progressed as smoothly as it had and completed ahead of the original schedule had it not been

because the Internet provided means for disseminating and sharing information in a timely manner.

9 A New Paradigm Of Discovery

Judging from the current prevailing trends in federal spending, healthcare and social reforms, it is very likely that bioinformation, disease database maintenance, intelligent software for extracting knowledge from these databases, will play a major role in the future of biomedicine. Disease diagnostics, prognostics and therapeutics will rely more on biodata, and bioinformation and bioknowledge derived therefrom, than on guesswork, chemistry or pharmacology.

Up until recently, successful therapeutic approaches target initial causative agents such as infectious microorganisms, or empirically target a single step of a multi-step complex disease process. Current therapeutic intervention, and therefore drug discovery efforts, are aimed at the molecular events of the disease process itself.

Conventional approaches focus on identifying, isolating, purifying targets; determining target sequence and three dimensional structures; applying rational drug design, molecular modeling for docking active sites; synthesizing, screening and evaluating chemical compounds for clinical test and FDA approval.

Traditional methods in molecular biology normally work on a “one gene in one experiment” basis, leading to a very limited throughput, and a global picture of the gene function is hard to obtain. In the past several years, the biochip technology has created a lot of interest and it promises to monitor the entire genome on a single chip so that researchers can have a better picture of the interactions among the thousands of genes in the genome simultaneously.

A biochip technological challenge involves achieving all the market-oriented parameters at a cost that supports a commercially acceptable price. Currently, DNA chips cost between $100 and $450 each. In a tight managed-care marketplace, that places a premium on technologies that can either show quick savings or more-efficient results. When the unit price for biochip reaches about $10, the biochips will be used more widely, and the biochip will be a commodity.

In healthcare, bioinformatics raises a number of future perspectives:

- If a target functions in a biological pathway, are there any undesirable effects from interactions of this pathway with associated pathways.
- Are there non-active sites that may yield greater specificity and this reduces side effects arising from interactions with structurally and evolutionarily related targets.
- The specificity, selectivity and efficacy of the small molecules.
- Time course of a disease process, i.e., a more dynamical study.
- Others.
Table 5. Healthcare in pre-genomic and post-genomic eras. (HAL thanks Dr. David Wang, Motorola BioChip Systems for).

<table>
<thead>
<tr>
<th>Pre-Genomic Era</th>
<th>Post-Genomic Era</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease description</td>
<td>Disease mechanism</td>
</tr>
<tr>
<td>Uniform disease</td>
<td>Disease heterogeneity</td>
</tr>
<tr>
<td>Patient homogeneity</td>
<td>Individual variability – SNP</td>
</tr>
<tr>
<td>Universal Rx strategy</td>
<td>Patient risk profiling - pharmacogenomics and targeted care</td>
</tr>
</tbody>
</table>

Though data and information are important for understanding diseases and other healthcare issues, for any data and information to be reliably useful, careful statistical analyses with sufficient samples are necessary. Otherwise, the derived conclusion and inferences may not be a true reflection of the studies. And in each regime of studies, a different approach may be employed. For example, for biochips are an appropriate platform for studying pharmacogenomics.

Figure 9. Samples sizes required for different purposes can be different. For example, the Human Genome Project produces 3 billion ($3\times10^9$) bases from the DNA of a few individual. There is a nucleotide difference in every 1000 nucleotide. Thus the SNP project has about $10^6$ bases involving a few scores of individuals. (Figure adapted from Dr. David Wang, Motorola BioChip Systems).

10 Historical Dé Jàvu?

The crux of hard reality is that if one has no vision and is too inflexible, one is permanently left behind.

The Agrarian Revolution started in Asia, lasted for tens of thousands of years, and ended in Europe. The Industrial Revolution started in Europe, lasted for about
two centuries (1760s-1950s), of which 90 years (1860s-1950s) were in United States. History has taught us two important lessons:

- With each successive revolution, the time span of the revolution is much shorter than that of its predecessor.
- A revolution that initiated in one region of the world does not guarantee that the revolution will end in that part of the world. Indeed, historically, a revolution commenced in one part of the world never ended in that same part of the world.

The twentieth century was very eventful: The First World War, The Great Depression, The Second World War… The first half of the century was shaped by the physical sciences but its second by biology. The first half brought the revolution in transportation, communication, mass production technology and the beginning of the computer age. In warfare, it brought in the nuclear weapons and the irreversible change in the nature of warfare and environment. The public’s preoccupation with the physical sciences and technologies, and the immense upheavals in the human condition meant that biology and medicine could only move to the center stage somewhat later, that is, in the second half.

The Information Revolution began in the United States two and a half decades ago. So did the biotechnology revolution. Will these revolutions reach their respective peaks in United States? This only time will tell.

Now we enter the twenty-first century, biology will likely continue to play a dominant role. First, biotechnology has shed its pristine academic garb and headed into the marketplace for more than two decades. Second, biotechnology has become so data-intensive and information-rich that it is essentially a part of information science. Third, the techniques of biotechnology have now been used for a new kind of war – biowarfare. Similarly, information technology has led to infowarfare. A new kind of war has just begun. Yet if we weigh the good and the bad, biotechnology and information technology advances have tended to improve life spans and prosperity.

---


* In October of 1999, when HAL presented a lecture on “Turbocharging bioinformation for drug discovery” at an international congress, the remark on the Agrarian, Industrial and Information Revolutions aroused quite a lot of interest.