CHAPTER VI:

INFORMATION TECHNOLOGY'S ROLE IN LIFE SCIENCES RESEARCH & DEVELOPMENT

By Cassandra Ingram and Gurmukh Gill*

Information technology (IT) has dramatically reduced the costs, increased the speed, and improved the productivity of life sciences research and development (R&D). Life sciences R&D, in turn, has opened up new challenges and opportunities for IT applications. This virtuous cycle has contributed to a whole new frontier for knowledge generation. For example, the confluence of IT and biological advances made possible the mapping of the entire human genome and genomes of many other organisms in just over a decade. These discoveries, along with current efforts to determine gene and protein functions, have improved our ability to understand the root causes of human, animal and plant diseases and find new cures. Furthermore, many future IT innovations will likely be spurred by the data and analysis demands of the life sciences.

This chapter describes the relationship between IT and life sciences R&D. We provide information on the life sciences market for IT goods and services. While currently the life sciences market comprises only a small part of the total IT market, it is a vibrant market and growing rapidly. Job opportunities that combine the skills of life and computer scientists are also expected to expand.

Bioinformatics Is Vital for Advances in the Life Sciences

In recent years, innovations where IT and the life sciences converge have created vast quantities of data. The development of automated DNA sequencing and other innovative methods have reduced the costs and time needed to discover the genetic makeup of various organisms.

* Ms. Ingram (cassandra.ingram@esa.doc.gov) is an economist and Mr. Gill (gurmukh.gill@esa.doc.gov) is senior executive for economic research in the Office of Policy Development, Office of the Chief Economist, Economics and Statistics Administration.

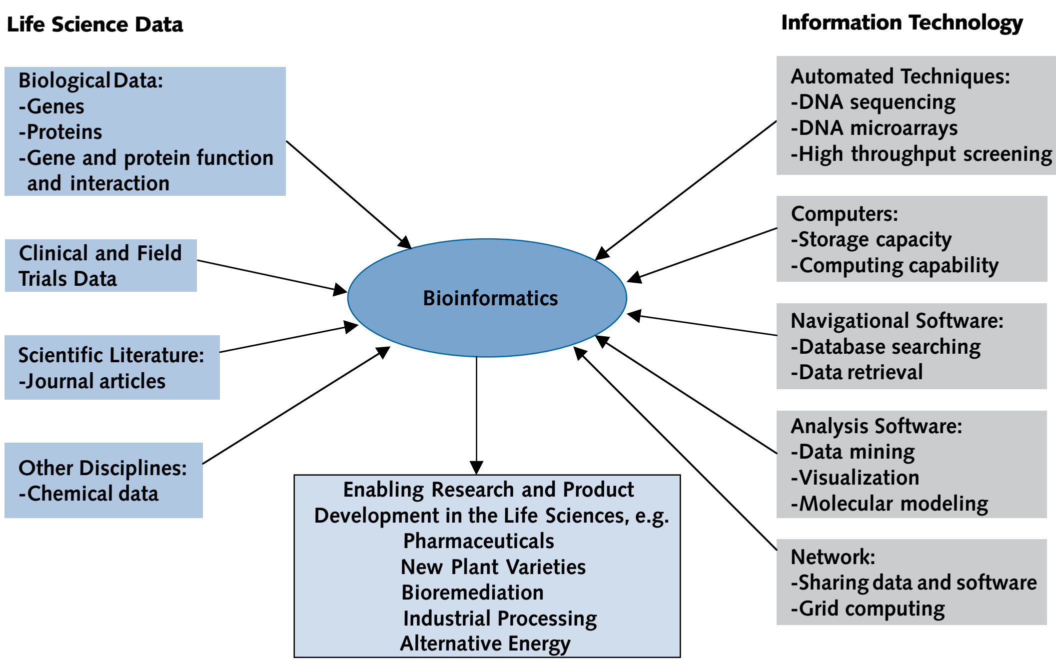
Additionally, innovations such as high-throughput screening and microarrays are enabling studies of gene and protein functions (i.e., genomics and proteomics).

The expanding complexity and diversity of information also pose new challenges. For example, data produced from studies of protein and gene function and interaction can require the consolidation of information in various formats and from diverse sources. Additionally, data may not be in a standard or readily usable format, particularly if they are not available electronically or are text-based information from scientific literature, patents and clinical trials. Critical biological information also is found throughout various scientific disciplines, such as the chemical sciences.[1] The ability of researchers to manage and analyze these diverse and extensive data is critical to the future success of life sciences R&D.

THE NEED FOR DATA MANAGEMENT AND ANALYTIC IT TOOLS

The needs described above have spawned the research field "bioinformatics,"[2] which focuses on the use of IT to collect, organize, store, interpret, share, and analyze biological data. Developments in bioinformatics will be critical for facilitating R&D in areas such as human and animal health, agriculture, industrial processing, natural resource recovery, and environmental remediation. Figure 6.1 illustrates the IT and life sciences components of bioinformatics.

Figure 6.1. Bioinformatics Uses Information Technology to Manage and Analyze Information Generated by the Life Sciences



---

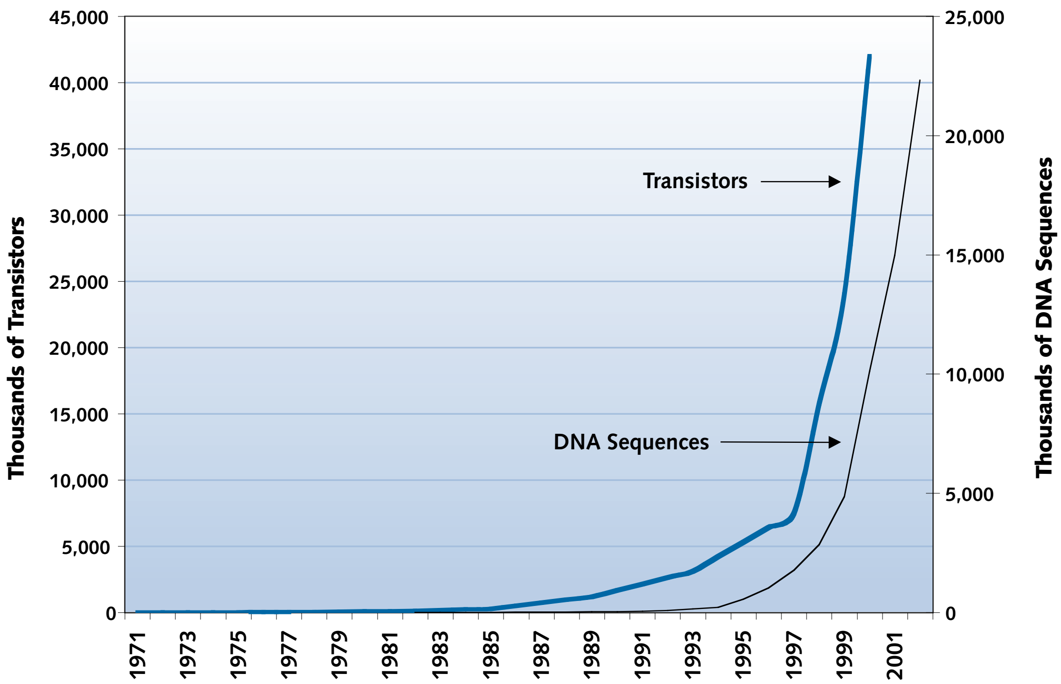
[1] For example, patient response to chemically-formulated pharmaceuticals.

[2] Computational biology research is the more traditional discipline that relates to bioinformatics, and consists of data analysis and interpretation and the development of new algorithms and statistics in the biological sciences.

Computing Power and Networks

Data storage, management, and analysis requirements in the life sciences are outpacing current computing capabilities, even though computing power continues to increase. Computing power has expanded at roughly the same rate as public DNA sequences located at the National Institutes of Health (NIH)—both have been doubling every 18 to 24 months for several years (Figure 6.2).[3] In 1971, only 2,250 transistors were on an integrated circuit; by 2002 there were 42 million. Similarly, only 606 DNA sequences were housed at NIH's Genbank in 1982. The number of sequences climbed to 22.3 million in just 20 years.

Figure 6.2. Number of Transistors per Integrated Circuit vs. DNA Sequences in Genbank



Source: Data on computing power, or the number of transistors per integrated circuit, is from Intel at http://www.intel.com/research/silicon/mooreslaw.htm. Data for the number of public DNA sequences of several organisms in Genbank is from the National Center for Biotechnology Information, National Institutes of Health at http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html.

Researchers are exploring strategies to increase computing capabilities for diverse life sciences purposes, such as computer-based drug design. One strategy is to expand the potential of supercomputers. Another is to unify computing resources through development of global, or grid, computing networks. These networks link the power of individual PCs or supercomputers.[4] Researchers have achieved teraflop speeds (a trillion floating point operations per second) using

---

[3] Mark S. Boguski, "Bioinformatics—A New Era," *Trends Guide to Bioinformatics* (Elsevier Science, 1998).

[4] See http://www.cs.mu.oz.au/~raj/papers/TheGrid.pdf (accessed Sept., 2003).

these grid networks. In the future they hope to achieve petaflop speeds (a thousand trillion floating point operations per second).[5]

The Internet is a particularly valuable tool for life sciences researchers, especially for those with limited resources. It enables researchers to tap into data, software, and computing power.

New software developments also promise life sciences researchers' the ability to acquire, format, search, and analyze disparate data sources and types. Data mining and visualization software[6] are improving scientists' ability to screen data and identify important relationships. One software analysis tool, 3-D molecular modeling, facilitates studies of the function and relationships between biological molecules. These analysis tools are enabling scientists to conduct some research at the computer instead of at the lab bench, reducing the time and costs of life sciences R&D.

Standardization enables researchers to link different databases and software programs, thereby making the discovery process more efficient. Standardizing scientific nomenclature, databases and software helps researchers exchange and analyze information more easily.[7] Researchers also benefit from the development of technical and computational standards for hardware and software.

Given the diversity of and rapid developments in life sciences databases and software, standardization is too difficult and costly for any organization to undertake alone. To overcome this obstacle, the Interoperable Informatics Infrastructure Consortium (I3C)[8] was founded in 2001 to collectively address some standardization problems. I3C is an international consortium that includes life science and IT participants from private industry, government institutions, academia and other research organizations. They develop and promote "global, vendor-neutral informatics solutions that improve data quality and accelerate the development of life science products." I3C's accomplishments include standards developed to identify and access biologically significant data and a method that simplifies data retrieval from multiple databases.

BIOINFORMATICS ASSISTS DRUG DISCOVERY AND DEVELOPMENT

Bioinformatics is improving the R&D process in drug discovery and development. IT tools have become important for managing and screening genetics data and for modeling outcomes in drug development. New developments in bioinformatics and genetics, such as pharmacogenetics (i.e., the study of the relationships between diseases, genes, proteins, and pharmaceuticals), will

[5] See whatis?com at http://whatis.techtarget.com/definition/0,,sid9_gci212778,00.html (accessed Sept., 2003).

[6] *Instrument Business Outlook.* Vol. 11, i12 (Sept 30, 2002) 3(1).

[7] *Bioinformatics Workshop.* A report produced for the Research and Resources Infrastructure Working Group, Subcommittee on Biotechnology, National Science and Technology Council, White House Office of Science and Technology Policy, and prepared by Tracor Systems Technologies, Inc. under contract with Krasnow Institute of Advanced Studies and George Mason University. (February, 1998).

[8] Information on I3C can be found on their website at http://www.i3c.org/ and http://www.i3c.org/demos/bio2002/i3cdatasheet.pdf (accessed Aug., 2003).

enable researchers to identify quickly a patient's genetic predisposition to contract certain diseases as well as their potential drug response.

Costs of drug discovery have been escalating, nearly tripling since 1991.[9] It now costs $900 million and takes 15 years on average to develop a new drug.[10] Clinical trials constitute most of these costs. Additionally, about 75 percent of drug development costs can be attributed to failures. Analysts anticipate that advances in IT and the biological sciences—such as the computer-enabled ability to quickly screen drug candidates and predict drug responses—could lower failure rates substantially, reducing costs by as much as one-third and time by as much as two years.

BIOINFORMATICS IS A KEY LIFE SCIENCES R&D ACTIVITY

IT tools and bioinformatics R&D are key to remaining competitive for biotechnology and pharmaceutical companies. These companies are expanding IT capabilities by developing in-house R&D programs in bioinformatics, acquiring bioinformatics companies, and partnering with IT companies, bioinformatics firms, and the public sector (e.g., the federal government and universities).

Survey of Firms Engaged in Biotechnology Activities

Results from a survey on the use of biotechnology in U.S. industry[11] demonstrate the importance of bioinformatics to life sciences firms. Table 6.1 displays some of the biotechnology activities of survey participants. About 30 percent of survey respondents indicated that bioinformatics was one of their biotechnology R&D activities. While 29 percent conducted bioinformatics research, only 3 percent reported having bioinformatics products or processes that were marketed or in production. Firms were also substantially engaged in activities highly dependent on bioinformatics such as genomics, DNA sequencing and synthesis, drug design and delivery, synthesis and sequencing of proteins and peptides, and combinatorial chemistry and 3-D molecular modeling.

Table 6.2 presents information on bioinformatics activities by application. Firms focusing on human health applications constituted the greatest number of survey respondents. Therefore, it is not surprising that 81 percent of the respondents (i.e., 247 out of 304) that conduct bioinformatics research were addressing human health problems. Within any application, between 19 to 45 percent of firms indicated that they conducted bioinformatics research.

---

[9] The Boston Consulting Group, *A Revolution in R&D: How Genomics and Genetics Are Transforming the BioPharmaceutical Industry*, (2001) at http://www.bcg.com and http://www.bcg.com/publications/files/eng_genomicsgenetics_rep_11_01.pdf (accessed Oct., 2003).

[10] Tufts Center for the Study of Drug Development at http://csdd.tufts.edu/ and http://csdd.tufts.edu/NewsEvents/RecentNews.asp?newsid=29 (accessed Oct., 2003).

[11] Survey data from *Critical Technology Assessment of Biotechnology in U.S. Industry*, U.S. Department of Commerce, Technology Administration and Bureau of Industry and Security (August 2002).

Table 6.1. Percent of Survey Respondents by Biotechnology Activity, 2002

	Conduct research on/in	Approved, marketed, or in production		Total
		Product(s)	Process(es)	
DNA-based				
Bioinformatics	29	2	1	30
Genomics, pharmacogenetics	29	3	2	30
DNA sequencing/synthesis/ amplification, genetic engineering	39	5	3	43
Biochemistry/Immunology				
Drug design & delivery	33	4	2	38
Synthesis/sequencing of proteins and peptides	27	3	1	30
Combinatorial chemistry, 3-D molecular modeling	18	1	0	19

Note: The total number of responses to the biotechnology activity question was 1021. Percents do not add up to 100 percent because firms can have more than one activity.

Source: Survey data from *Critical Technology Assessment of Biotechnology in U.S. Industry*, U.S. Department of Commerce, Technology Administration and Bureau of Industry and Security, August 2002.

Table 6.2. Number of Survey Respondents Indicating Bioinformatics Research Activities by Application, 2002

Application	Number of firms in application	Conduct bioinformatics research
Human Health	780	247
Animal Health	144	37
Agricultural & Aquacultural/Marine	128	41
Marine & Terrestrial Microbial	41	19
Industrial and Agricultural-Derived Processing	132	45
Environmental Remediation and Natural Resource Recovery	41	12
Other	160	30

Note: The total number of firms that responded to the biotechnology survey was 1,031, and 304 of these firms indicated that they had some activity in bioinformatics. The number of firms by biotechnology application does not add up to the total number of firms that responded to the survey because firms were classified in an application if they indicated it as either a "primary" or "secondary" focus.

Source: Survey data from *Critical Technology Assessment of Biotechnology in U.S. Industry*, U.S. Department of Commerce, Technology Administration and Bureau of Industry and Security, August 2002.

PUBLIC INSTITUTIONS SUPPORT BIOINFORMATICS R&D

Public institutions have made significant investments in research areas where the life sciences and IT converge. They supported and organized endeavors to map the genomes of several organisms, created databases, developed data retrieval and analysis software, conducted IT and other bioinformatics-related R&D, and funded bioinformatics training programs. These activities help drive the demand for IT goods and services in the life sciences.

The National Institutes of Health (NIH) represents a majority of Federal spending for life sciences R&D, which is focused on human health. A five-year campaign to double NIH R&D funding to $27.3 billion was completed in 2003.

Several NIH programs and initiatives that are closely linked with bioinformatics R&D, such as the Human Genome Project,[12] will get funding boosts in 2004.[13] Also in 2004, NIH plans to devote $35 million for developing new life science approaches and technologies, such as bioinformatics. The National Institute of Allergy and Infectious Diseases, which is the lead institute in bioterrorism research, will receive increased funding. Bioinformatics is a key element in developing drug and vaccine candidates as countermeasures to potential bioterrorism pathogens. NIH also makes sizable investments in IT R&D. In 2003, NIH allocated $336 million for information technology R&D, representing about 1.3 percent of their $23.6 billion R&D budget.

Other federal agencies, such as the National Science Foundation (NSF) and the Department of Energy (DOE), also support life sciences and IT R&D, especially in disciplines other than human health (e.g., environmental biology, the plant sciences and alternative energy sources).[14] A little over 10 percent of NSF's $5 billion budget and 2 percent of DOE's $22 billion budget are devoted to biological R&D. NSF's biological sciences R&D budget in 2004 shows a 20 percent increase (to $82 million) in emerging frontiers research, which includes bioinformatics. Of this, information technology research for the biological sciences would rise 10.3 percent to $7.5 million. DOE is a major supporter of advanced scientific computing research. R&D spending for this program is expected to increase 4.2 percent, to $173.5 million.

IT Market and Job Opportunities in the Life Sciences

The life sciences are opening up a new frontier for profitable IT innovations and applications. Although the life sciences represent only a small fraction of the entire IT market, that portion is growing at a substantial rate. A number of market research reports describe the life science market for IT in the United States and globally as dynamic with vast growth potential.[15] Global

---

[12] The Human Genome Project funding will increase 2.8 percent, from $465 million in 2003 to $478 million in 2004.

[13] See "Summary of the FY2004 President' Budget" for NIH at http://www.nih.gov/news/budgetfy2004/fy2004presidentsbudget.pdf (accessed Sept., 2003) and NIH budget highlights from the American Association for the Advancement of Science at http://www.aaas.org/spp/rd/nih04p.pdf.

[14] See NSF's FY 2004 budget request at http://www.nsf.gov/bfa/bud/fy2004/toc.htm and budget request for the biological sciences http://www.nsf.gov/bio/bio_bdg04/bionarr04.htm (accessed Sept., 2003). See DOE's FY 2004 budget request at http://www.mbe.doe.gov/budget/04budget/content/highlite/highlite.pdf (accessed Oct., 2003). DOE also has contributed to Human Genome research and has a new Genomes to Life program, which is focused on the "function and control of molecular machines for energy and environmental applications."

[15] The four primary market research firms that estimate the bioinformatics market are the International Data Corporation (http://www.idc.com/getdoc.jhtml?containerId=TEA001604), Frost & Sullivan (http://www.frost.com/prod/servlet/frost-home.pag), Strategic Directions International (http://www.strategic-directions.com/), and Front Line Strategic Consulting (http://www.frontlinesmc.com/).

bioinformatics market estimates vary widely, ranging from $0.8 billion to $14.6 billion in 2002.[16]

The life sciences market for IT is small relative to the whole IT market. The most optimistic estimate of the U.S. life science market for IT was $7.4 billion[17] in 2002. This was less than 0.9 percent of the value added of the IT-producing industries (which was $829 billion in 2002).[18]

However, the life sciences market segment appears to have strong growth potential. During the economic downturn of 2001, growth in value added of IT producing industries declined by 5.6 percent (see chapter 1). Market research firms reported that during this same period growth rates in the IT life science market were high overall. They also expect growth rates to continue to rise, ranging from 19 to 25 percent annually[19] until at least 2006.

JOB OPPORTUNITIES IN BIOINFORMATICS

Increasingly companies and research organizations are seeking workers with more formalized training that have the skills of both computer and life scientists. The high starting salaries of bioinformaticians, $65,000 to $90,000 per year, reflect the strong demand for bioinformatics employees.[20]

Computer specialists[21] in the life sciences are among the high-technology employment categories currently experiencing job growth. Occupational projections by the Bureau of Labor Statistics[22] (BLS) suggest that employment of computer specialists in the Drug and Research &

[16] The large discrepancy in market and growth estimates mostly depends on market segment covered and organizations included, the years spanned, and assumptions made related to future market projections.

[17] The International Data Corporation estimated that the IT life science market was about $14.6 billion worldwide and that the U.S. market represented 51 percent of this total.

[18] The estimate of 0.9 percent is based on total revenue and is not value added; i.e., where purchased inputs and labor costs from outside the firm are excluded. See chapter 1 on information technology producing industries.

[19] Growth rates presented in the market research reports were compound annual growth rates.

[20] Stephan, Paula E. and Grant Black, "Bioinformatics: Emerging Opportunities and Emerging Gaps," *Capitalizing on New Needs and New Opportunities in Biotechnology and Information Technologies: Government-Industry Partnerships in Biotechnology and Information Technologies*, ed. Charles W. Wessner., Board on Science, Technology and Economic Policy, National Research Council (Washington, DC: National Academy Press, 2001).

[21] The occupations included in the BLS aggregate "Computer Specialists" are computer programmers, computer and information scientists (research), computer systems analysts, computer software engineers (applications), computer software engineers (systems software), computer support specialists, database administrators, network and computer systems administrators, network systems and data communications analysts, and all other computer specialists.

[22] See the following BLS websites for data and analysis http://www.bls.gov/oco/cg/cgs009.htm, http://ftp://ftp.bls.gov/pub/special.requests/ep/ind-occ.matrix/ind_pdf/i808730.pdf, and http://ftp://ftp.bls.gov/pub/special.requests/ep/ind-occ.matrix/ind_pdf/i422830.pdf (accessed Sept., 2003).

Testing Services industries[23] will continue to grow rapidly for several years. BLS estimates that Drug industries will increase computer specialist jobs by 60 percent, from 5,545 in 2000 to 8,859 in 2010. BLS expects computer specialist jobs in Research & Testing Services industries to increase 68 percent, from 42,567 in 2000 to 71,549 in 2010.

A survey of biotechnology use in U.S. industries further demonstrates the demand for computer specialists in the life sciences.[24] While computer specialists only accounted for 6.2 percent of the respondents' biotechnology R&D workforce, growth in this occupational category was the fastest among the biotechnology R&D job categories. Between 2000 and 2002, responding companies added 1,236 computer specialist jobs, a 20 percent increase.

Conclusion

Researchers have become increasingly reliant on IT tools to reduce the costs and boost the productivity of life sciences R&D. At the moment, the life sciences market for IT goods and services is small. However, the life sciences market for IT remained strong even during the 2001 recession and is expected to grow rapidly. Bioinformatics employment opportunities also have been expanding.

[23] While the Drug and Research & Testing industries are not entirely composed of life sciences industries, a study entitled *Critical Technology Assessment of Biotechnology in U.S. Industry*, U.S. Department of Commerce, Technology Administration and Bureau of Industry and Security (October 2003), (http://www.technology.gov/reports/Biotechnology/CD120a_0310.pdf) found that firms engaged in biotechnology activities mostly were classified under one of two industries, either Drugs or Research & Testing Services, using the Standard Industrialized Codes. These two industries were used in this analysis as the most representative of the life sciences industries.

[24] U.S. Department of Commerce, Technology Administration and Bureau of Industry and Security, *Critical Technology Assessment of Biotechnology in U.S. Industry* (October 2003), (http://www.technology.gov/reports/Biotechnology/CD120a_0310.pdf).