Proper Citation:
Hwa A. Lim, “Bioinformatics and cheminformatics in the drug discovery cycle”, In: Lecture Notes in Computer Science 1278, Bioinformatics, Ralf Hofestädt, Thomas Lengauer, Markus Löffler, and Dietmer Schomburg (eds.), (Springer-Verlag, Berlin, 1997), pp. 30–43.

Bioinformatics and Cheminformatics in the Drug Discovery Cycle

by

Hwa A. Lim, Ph.D., MBA
D'Trends, Inc.
3521 Ryder Street
Santa Clara, California 95051-
USA
hal@d-trends.com
http://www.d-trends.com
(February 1997)

0. Abstract

This is a slightly modified version of a report presented at a workshop of the GCB'96 Conference. We describe the paradigms of bioinformation and cheminformation. The rise of bioinformatics, a new subject area that has been receiving a lot of attention in recent months, is also chronicled. The dynamics forcing pharmaceutical companies to undertake major infrastructure investments in new, complex and very data-intensive drug discovery technologies are discussed, and the roles of bioinformatics and cheminformatics in the context of drug discovery are also given.

Keywords: bioinformatics, computers, database, disease, drug, genome research, sequencing.

1. Introduction

The prevailing view in this post-Cold War era is that biology has jostled to the center stage at the expense of the physical sciences. This is a fallacy.

In these remaining centennial years, 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 about revolutions 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, and pinnacled with the moon shot. 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, discovery of the structure of DNA, early harbingers of the green revolution are all proud achievements [1]. 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 technology, nuclear magnetic resonance, ultrasound technique, PCR, X-ray crystallography, 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 [2].

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 data that are being generated as a result of the international genome project [3]-- a project that has been hailed as the "moon-shot" of biology - and provide the teraflop compute power required for complicated analyses to penetrate the deepest secrets of biology. Consequently, the time is ripe for a marriage made in heaven between biology and computer science-- biocomputing; and the study of information content and information flow in biology and chemistry, i.e., bioinformatics and cheminformatics, respectively.

2. The Rise of Bioinformatics

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: 1) our knowledge of biology has exploded in such a way that we need powerful tools to organize the knowledge itself; 2) the questions we are asking of biological 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 systems and processes." It has evolved to serve as a bridge between the observations (data) in diverse biologically-related disciplines and the derivations of the understanding (information) about how the systems or processes function, or in the case of a disease, dysfunctions and subsequently the application (knowledge), or in the case of a disease, therapeutics (See, for example, `http://www.awod.com/netsci/`).

Cheminformatics, which came after bioinformatics, is defined in an analogous manner.

2.1. 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 [4]. 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, and with help from Professor Charles R. Cantor (then Chairman of the College of Physicians & Surgeons at Columbia University) and Professor Jospeh E. Lannutti (then Director of Supercomputer Computations Research Institute at Florida State University), the author convened the very first conference in bioinformatics. *The First International Conference on Electrophoresis, Supercomputing, and The Human Genome* was held at the Florida State Conference Center, Tallahassee, April 10-13, 1990. Though the title did not contain the word "bioinformatics", bioinformatics was a major part of the conference. Among the more prominent 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 (then Director, Lawrence Livermore National Laboratory Genome Program), Temple Smith (then Director at Dana Farber Cancer Center of Harvard Medical School), Alexandar Bayev (then Chairman, USSR Genome Program), Boris Kaloshin (USSR Dept. of Sc. & Tech), M. Durand (French Embassy), N. Shimizu (Head, Department of Molecular Biology, Keio University School of Medicine), I. Endo (RIKEN, Japan), N. Nord{\'e}n (Sweden), and others (120 participants in total). The conference was funded by The US Department of Energy, and The Florida Technology Research and Development Authority, Thinking Machines Corp., Digital Equipment Corp., CRAY Research Inc. A proceeding volume was compiled [5]. Note that the sponsors were primarily federal and state agencies, and general-purpose computer companies.

2.2. 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, June 4-7, 1992. This conference was originally planned for St. Petersburg (Leningrad), USSR. The breakup of the Former Soviet Union forced the author 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 is 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., The Technological Research & Development Authority, The US Department of Energy, The US National Science Foundation. A second proceeding volume was edited [6] to bring the subject area to the then relatively small community. Note the participation of federal and state agencies, special-purpose computer companies and publishing houses.

The third conference, *The Third International Conference on Bioinformatics & Genome Research*, took place at the Augustus Turnbull III Florida State Conference Center, Tallahasee, Florida, June 1-4, 1994. It was partially funded by Compugen Ltd., Eli Lilly and Company, MasPar Computer Corp., World Scientific Publishing Co., Pergamon Press, The US Department of Energy, The US National Science Foundation, The US National Institutes of Health, The International Science Foundation. The proceedings were gathered in a volume [7]. A noteworthy point is that the sponsors were federal, state and international agencies, special-purpose computer companies, pharmaceutical companies and publishing houses.

2.3. Bioinformatics Conference Going Commercial and Online

This biennial conference series was taken over by CHI (http://www.healthtech.com} in 1994. Due to the popularity of the subject area, CHI decided to make the conference series an annual event. *The Fourth International Conference on Bioinformatics & Genome Research* was held at Hotel Nikko, San Francisco, June 5-7, 1995. There was no conference proceedings for this year because of complications with copyrights. *The Fifth International Conference on Bioinformatics & Genome Research* just took place at the Baltimore Inner Harbor Hotel from June 10-11, 1996. Some of the papers presented were published in Gene-Combis (an online publication). The upcoming *Sixth International Conference on Bioinformatics & Genome Research* will be held at The Fairmont Hotel, San Francisco, June 11-12, 1997.

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.

Among the first international teleconferences was that held in 1992 by Global University in the USA, a Divisional Activity of Global Systems Analysis and Simulation Association in the USA (GLOSAS/USA)(http://www.wiu.edu/users/milibo/wiu/resource/glosas/cont.htm), in which the author took part. Credit for the first teleconference in biologically related work goes *to Intelligent Systems in Molecular Biology*, held in 1994.

2.4. Related Publications and Conferences

To do justice to the area, the following related books [8--14] (Ref. 9 is decidedly the first of its kind, which talks about information content in biological systems. The book is a collection of articles presented at *The Symposium on Information Theory in Biology*, organized in Gatlinburg, Tennessee, Oct 29-31, 1956. must be cited. This list is by no means exhaustive. There are also many related conferences, workshops and meetings. Among them are *Intelligent Systems in Molecular Biology*; Hilton Head Meeting; *The World Congress on Computational Medicine, Public Health and Biotechnology*; *The German Conference on Bioinformatics*; *Integrative Approaches to Molecular Biology*; and many others. Many computer, mathematics and statistics conferences are also beginning to include sessions on bioinformatics or biocomputing (See for example, *The ACM International Conference on Supercomputing*, and *The International Conference on Mathematical and Computer Modelling and Scientific Computing*.

It now seems that *The Bioinformatics & Genome Research conference* series will continue for many years to come. *The Intelligent Systems in Molecular Biology Conference* series is also doing extremely well and will probably last for a long time.

Lest we forget, we must also mention the impressive bioinformatics activities along the Pacific Rim (See for example, `http://biomed.nus.sg/biocomp/`; `http://life.anu.edu.au/`) and in Europe (See for example, http://www.embl-heidelberg.de/, http://www.genethon.fr/; `http://www.ebi.ac.uk/`). Even though the US initiated bioinformatics and the German bioinformatics effort started a few years later in 1993, the Germans seem to have done quite a lot for the subject area. Currently, the German government has committed \$16,000,000 for the project. The recent *First International German Conference on Bioinformatics* [14] went off to an excellent start. There is every indication that it will last for a long time to come.

On May 3rd, 1996, a BioMASS panel, in conjunction with the BioScience Career Fair and sponsored by AAAS-- publisher of the Science magazine, was held at Stanford University Medical Center. The author took part as a bioinformatics panelist. Subsequently, a series of articles and interviews appeared in the Science magazine [15] (`http://www.aaas.org/`) "Bioinformatics" became a buzzword soon afterwards.

3. Genomic Companies As Service-Oriented Companies

Let us now turn to benchwork briefly. Many genomics companies and centers have unique, high-throughput, cost effective technology to do sequencing and to collect data. But, as shown in Table~1, data is not "commercializable", but information is. This leads naturally to a conceptual flowchart of biodata, as depicted in Figure~1. Or in terms of physical design, the corresponding databases as illustrated in Figure~2.

A table to compare and contrast *data* and *information*.

Data are...	**Information is…**
Stored facts	Presented facts
Inactive (they exist)	Active (enables doing)
Technology-based	Business-based
Gathered from various sources	Transformed from data

Biodata

↓

Bioinformation

↓

Bioknowledge

↓

Next generation genomics/drug discovery

A flowchart to show the paradigm of ***biodata***. The prefix "bio" can equally be substituted for "chem".

Database

↓

Infobase

↓

Knowledgebase

↓

Disease treatment

The paradigms of ***biodata*** and ***chemdata*** presented in a more physical form, i.e., as various databases.

Figure~3 shows that bioinformatics drives the decision making process by:

1. supporting large scale sequencing, utilizing proprietary, high throughput sequencing technology,
2. incorporating sequencing-derived data such as clone signatures, genes, etc,
3. maintaining and operating a unique database and knowledgebase.

In order to maintain such a scheme, a possible strategic plan is outlined in Table~2 [16,17].

High Throughput Sequencing (Screening) Technology

+

Bioinformatics (Cheminformatics)

=

GeneDatabase/KnowledgeBase

↓

$$Commercialization$$

A flowchart depicting the current sequencing and screening technologies to commercialization via bioinformatics and cheminformatics, respectively.

A chart showing the flow and planning of information, in particular, bioinformation. The sequence is: assessment, strategy and execution.

Assessment	**Strategy**	**Execution**
Current position	Future position	Adjust implementation Programs
Positional analysis	Objectives & goals	Carry out projects to Attain objectives & goals
Directives, assumptions	Change management plan	
Conclusions	Commitment plan	
	Strategic moves	

4. Drug Discovery

We shall now turn to drug discovery and see the role informatics plays.

4.1. The Drug Discovery Cycle and Informatics

We shall take as an example protease, which is a *raison d'etre* of many start-up pharmaceutical companies, such as Arris Pharmaceutical Corp. (http://www.arris.com/). Proteases are naturally occurring regulatory enzymes that break down proteins. They are found throughout the body and play a role in many human diseases: In the best-known case, the AIDS virus uses a protease to dismantle healthy proteins and uses them to build new viruses; in the case of the inflammatory disease asthma, a form of serine protease, tryptase, stimulates the production of chemicals such as histamine, which may cause asthmatic attacks; in osteoporosis, osteoclast cells attach to the surface of a bone and release a protease, Cathepsin K, which under certain conditions, eats away the bone and thus causing the disease; in yet another example, protease Factor Xa, Factor VIIa and Thrombin, that contribute to the formation of blood clots at the site of a damaged blood vessel, run amok leading to thrombosis, a form of clotting. Protease also plays a critical role in reproduction - the head of every sperm cell is packed with a protease which the sperm uses to chew through the wall of the egg to complete fertilization.

In this particular case of protease, like in most other cases, drugs are usually designed to inhibit protease actions. The biggest hurdle in developing protease inhibitors, however, is that proteases are so omnipotent. Thus side effects can be overwhelming unless the drugs are very specific.

Usually, drugs are only developed when a particular biological target for that drug's action has already been identified and well studied, such as the case of proteases. Until recently, drug development was restricted to a small fraction of possible targets since the majority of human genes were unknown. The number of potential targets for drug development is increasing dramatically, due mainly to the genome project [3]. Drug developers are presented with an unaccustomed luxury of choice as more genes are identified and the drug discovery cycle becomes more data-intensive. However, such choice requires that additional information about each of the genes be obtained so that the best target can be selected.

Bioinformatics, in the drug development context, aims to facilitate the selection of drug targets by acquiring and presenting all available information to the drug developers. The constant growth in available information (information content) requires implementation of a dynamic process (information flow) to ensure that the presented information is complete and up to date (See for example, http://www.basefour.com/what_is.html).

4.2. The Economics of Drug Discovery

Let us turn to the economics of the drug discovery cycle. Of the about 5,000 - 10,000 compounds studied, only one drug gets onto 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 (This is the 1994 figure. It is estimated that the corresponding figure in 1997 is of the order of $400 million) for each drug put onto the market for consumers [18]. The time required for approval is equally long, as shown in Figure~4. These phases constitute parts of the manufacturing,regulatory and cost factors of drug discovery.

Preclinical Testing (~3.5 years)
↓
Investigational New Drug Application
↓
Clinical Trials, Phase I (~1.0 year)
↓
Clinical Trials, Phase II (~2.0 years)
↓
Clinical Trials, Phase III (~3.0 years)
↓
New Drug Application (~2.5 years)
↓
Approval

The long and expensive procedure for gaining FDA approval of a pharmaceutical product.

Besides the long and expensive drug discovery cycle, other factors contribute to the rapidly changing landscape of drug discovery environment:

- advances in molecular biology and high throughput sequencing;
- demand fundamentals
  a. aging population of the baby-boomers,
  b. consumer demand for quality healthcare,
  c. expanded access and universal healthcare,
  d. new breakthrough technologies,
  e. consumer awareness of the quality of nutrition and supplements, and
  f. others; and
- supply fundamentals, among many others -
  a. hospital downsizing,
  b. insurers' reluctance to pay high reimbursements,
  c. transition to outpatient procedures,
  d. disease management,
  e. global managing, and
  f. others.

Due to these factors - regulatory, cost-effectiveness of drug discovery and the supply and demand fundamentals - the process of drug discovery is undergoing a complete overhaul. Consequently, companies, which have been reaping a fortune from the sales of drugs are expected to shift their focus to tap into information. A case in point is managed healthcare. In the managed healthcare treatment of cancer, for example, the federal government might limit treatments to two per patient, instead of the age-old "physicians shall do whatever it takes" - the Hippocratic Oath. For instance, a patient will be given chemotherapy, and then an operation, if necessary. If this still does not help, that will be it.

Thus, companies which maintain good databases for diseases will be able to, via some intelligent software or otherwise, predict the best course treatment for individual patients depending on the ethnic background, progression and stage of illness, age, sex, previous history and others. Or that they can tap into bioinformation and cheminformation to shorten the cycle of drug discovery, and thus making drug discovery more cost-effective.

5. 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. The general consensus is that future pharmaceutical discoveries will stem from biological information. Major pharmaceutical companies develop new core products. These companies are either slower in response; or they do not want to develop sequencing expertise nor maintain proprietary database in-house; or they do not want to commit the financial resources for such purposes. But they do want to respond quickly and do need access to comprehensive genetic, biological and chemical information for timely and accurate decision making.

Modern drug discovery, on the other hand, has been transformed by the industrialization and automation of research. The resulting explosion in the quantity and complexity of biological, chemical, and experimental data has overwhelmed the ability of the drug discovery industry to make sense of it. The data explosion, combined with the pressure to reduce costs and speed up drug discovery cycles, provides a strong demand for software and information products. Informatics integration is the key to unleashing the potential of modern drug discovery.

Increasing reliance on genomic information about disease targets and on chemical information is creating a data-oriented research environment in which collaboration among molecular biologists, molecular modelers, drug chemists and computer scientists is essential for efficient drug discovery. These disciplines are loosely coupled by computational science. The role of bioinformatics and cheminformatics has changed from a specialist niche tool to that of an essential corporate technology. The scope has also accordingly widened from a laboratory-based tool to an integrated corporate infrastructure. Indeed, biology has become so data-intensive that the whole scenario has been paralleled to what happened to physics some fifty years ago.

The technology is coming to fruition at a pace that outstrips the capacity of the current methodologies of managing and analyzing biological and chemical data. Genomics, combinatorial chemistry and high-throughput screening are recognized as the triumvirate of the new order of drug discovery.

Thus we are seeing bioinformatics divisions springing up in all major pharmaceutical companies to either partake in this exciting new area, or to partner with smaller, more nimble companies. Because of this, smaller companies are constantly being formed to take advantage of the window of opportunities, some of which survived, and many more of which floundered. In general, these small companies try to develop technologies, be it laboratory-based or information-based, 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.

As with any business, one has to be on the *qui vive* for quacksalvers. There are many companies out there trying to sell unproven technologies and many eager investors are misled into empty promises. For example, a small biotechnology company may claim to have a core technology to do high throughput sequencing. More often than not, the company also uses a complementary and more proven technology, for example, an ABI machine, as a control. However, it will have no qualms in presenting results from the complementary technology as results from the core technology when the unproven core technology fails to live up to expectations. Or somehow by a legerdemain of skillful massaging selected data to make them look convincing; or to put up a Potemkin village with heavy machinery of moving parts, computers of blinking lights, foyers of chandeliers, offices of mahogany executive desks, etc, redolent of achievements, successes and wealth. In other words, the turpitude of code of business ethics is redefined. Ultimately the

stakeholders, which include investors, tax payers, clients, employees, to name a few, are the ones to lose while a selected few reap in huge profits. Another pitfall is duplication of efforts, which can be quite bootless. For example, in cDNA sequencing, several companies are using different core technologies to sequence many of the same tissues when the resources can be better utilized to sequence other tissues. There are even instances in which companies do so just to prove the "higher" throughputness of their core technologies. The bottomline is once the data has been obtained, no one really cares how it was obtained, or by which technology!

6. Bioinformatics & Cheminformatics - Mission and Goals

Based on our earlier discussion of the future of pharmaceutical discoveries, a typical goal and mission of a bioinformatics or a cheminformatics division might include, among many other possibilities and combinations: 1) enabling corporate partners to accelerate identification of genetic information for gene-based drug targets; 2) validating this selection through sequencing-derived drug-genome interaction studies; 3) performing decision making by centering around intelligent interpretation of existing genetic information; 4) identifying what information may yet be needed, define what may yet be done; 5) packaging this information for efficient decision making throughout a partner's product development cycle.
The goals and mission may vary in accordance with local needs, and very much driven by applications and clients.

7. Bioinfobahn

Since bioinformatics is a marriage of computer and biology, it is not surprising that it is well kept abreast with advances in computer technology, in particular, the internet technology.
The internet came into being about twenty years ago as a successor to ARPANET, a US 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, information and knowledge disseminate at a much more timely rate. There are countless electronic publications on the net, as is obvious from the cited footnotes of this text. These publications appear in the form of regular ascii text, postscript, hypertext, Java and other derivations therefrom.
A good example of a biotech company that fully utilizes the internet technology is D'Trends, Inc. (http://www.d-trends.com). D'Trends, Inc. develops and sells proprietary software products and information technologies that drive modern drug discovery process. These products and technologies integrate and automate the full range of pharmaceutical business-critical processes to provide unprecedented levels of productivity. Employing advanced informatics centered around client/server technology and internet/intranet database development, D'Trends has established a name throughout the biopharmaceutical industry as a leader in drug discovery informatics.
An impressive example from the public sector is GenomeNet (http://www.genome.ad.jp/). GenomeNet is a Japanese computer network for genome research and related research areas in molecular and cellular biology. GenomeNet was established in 1991 under the Human Genome

Program (HGP) of the Ministry of Education, Science, Sports and Culture (MESSC). It provides public access services for database retrieval and analysis.
The counterpart in Germany is Gesellschaft für Biotechnologische Forschung mbH (GBF) (http://rzinet.gbf-braunschweig.de/). GBF was founded in 1976 as a spin-off of its forerunner the Gesellschaft für Molekularbiologische Forschung mbH (GMBF). It is financed by the Federal Ministry of Research and Technology (BMBF) and the State of Lower Saxony. GBF is characterized by long term projects for protecting the environment, and for dealing with the knowledge, diagnosis therapy and prophylaxis of diseases.

8. Discussions and Conclusion

Judging from the current prevailing trends in federal spending, healthcare and social reforms, and other *force majeure*, it is very likely that information, disease database maintenance, and intelligent software for extracting knowledge from these databases, will play a major role in the future of disease treatment. Disease therapeutics will rely more on data, and information and knowledge derived therefrom, than on guess work, chemistry or pharmacology.
Current successful therapeutics target initial causative agents such as infectious microorganisms, or empirically target a single step of a multi-step complex disease process. Therapeutic intervention, and therefore drug discovery efforts, should be aimed at the molecular events of the disease process itself. Currently, there are a number of technological limitations: 1) slow rate of cDNA sequencing; 2) high cost of sequencing; 3) poor quantification and incomplete representation of cellular mRNA, among others. While many companies and research centers are developing high throughput, cost-effective technologies, the focus downstream should be on data, and information and knowledge derived therefrom, rather than on guesses.
Thus, from a more technical point of view, drugs of tomorrow are somewhere in the vast and growing sets of data available. The market for drug discovery informatics presents an unprecedented opportunity to create value in the management and extraction of data and its conversion to information and knowledge. While the computer can never completely substitute for laboratory work, it can however minimize bench-work and thus making drug discovery more cost-effective. The ultimate goal is to hasten the coming of age of "desk-top drug discovery" by developing the operating system of choice for drug discovery and development. In this sense, many software companies are functioning as lableless pharmaceutical companies. As an example, the "The linguae francae discovery trade" of D'Trends (http://www.d-trends.com) unites 1) automated genomics database analysis for drug target site selection; 2) chemical information database analysis and large scale combinatorial chemistry project management; and 3) high-throughput screening project management for drug lead efficacy analysis. These integrated elements forge a connection between the drugs of tomorrow, and the vast amounts of proprietary and published data available to researchers today. The "linguae francae" is also flexible enough to accommodate all commonly used database engines (Sybase, Oracle and Illustra) and all versions of Unix. In addition, new data formats, databases, algorithms and analysis paradigms are readily absorbed into the automated workflow without major software modifications. The popular webbrowser "Netscape Navigator" provides friendly user interfaces from PC, Macintosh, and Unix workstations.
From a more biochemical point of view, 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. Bioinformatics raises a number of future perspectives: 1) if the target functions in a biological pathway, are there any undesirable effects from interactions of this pathway with associated pathways; 2) are there nonactive sites which may yield greater specificity and this reduces side effects arising from interactions with structurally and evolutionarily related targets; 3) the specificity, selectivity and efficacy of the small molecules; 4) time course of a disease process, i.e., a more dynamical study; and 5) others.

The crux of hard reality is that if one has no vision and is too inflexible, one is permanently left behind. Time and tide wait for no one in the exciting and vibrant field of informatics. More and more, not only in the drug discovery business, but also in other businesses, companies are built on process knowledge that controls production and product development systems, proprietary software, and ways of integrating and outsourcing complex pieces of a value chain - pieces that may reside anywhere or in different disciplines. The name of the game is "customization"; these days almost nobody is making money from "commoditized" products. But knowledge assets are the least stable part of any business. They are easily copied, or recruited away, or superseded by yet newer technologies. Indeed the primacy of knowledge assets means that companies can get in and out of business much more quickly than ever before [19], but *"ships in harbor are safe, but that is not what ships are built for!*"

9. Acknowledgements

The author would like to thank B. Hauser, J. Schmutz, and G. Varga for reading and editing the original draft.

10. Bibliography

1. Ochoa, G., and Corey, M.: The Timeline Book of Science, (Stonesong Press, Ballantine Books, New York, 1995).
2. Naisbitt, J., and Aburdene, P.: Megatrends 2000: Ten New Directions for the 1990s, (Avon Books, New York, 1990).
3. Mapping and Sequencing the Human Genome, (National Research Council, National Academy Press, Washington, D.C., 1988).
4. Bell, G.I., and Marr, T.G., (eds.): Computers and DNA, (Addison-Wesley Publishing Co., Redwood City, 1990).
5. Cantor, C.R., and Lim, H.A. (eds.): Electrophoresis, Supercomputing and The Human Genome, (World Scientific Publishing Co. (URL: {\tt http://www.wspc.co.uk}), New Jersey, 1991).
6. Lim, H.A., Fickett, J.W., Cantor, C.R., and Robbins, R.J. (eds.): Bioinformatics, Supercomputing and Complex Genome Analysis, (World Scientific Publishing Co., New Jersey, 1993).
7. Lim, H.A., and Cantor, C.R. (eds.): Bioinformatics \& Genome Research, (World Scientific Publishing Co., New Jersey, 1995).
8. Yockey, H.P. (ed.): Symposium on Information Theory in Biology, (Pergamon Press, New York, 1958).
9. Hunter, L., Searls, D., and Shavlik, J. (eds.): Proceedings of The First International Conference on Intelligent Systems for Molecular Biology, (AAAI Press, Menlo Park, 1993).

10. Smith, D.W. (ed.): BIOCOMPUTING: Informatics and Genome Projects, (Academic Press, New York, 1994).
11. Schomburg, D., and Lessel, U. (eds.): Bioinformatics: From Nucleic Acids and Proteins to Cell Metabolism, (VCH Publishers, Inc., New York, 1995).
12. Hofestädt, R., Kruckeberg, F., and Lengauer, T.(eds.): Infomatik in den Biowissenschuften , (Springer-Verlag, Heidelberg, 1993).
13. Collado-Vides, J., Magasnik, B., and Smith, T.F. (eds.): Integrative Approaches to Molecular Biology, (MIT Press, Cambridge, 1996).
14. Hofestädt, R., Lengauer, T., L{\"o}ffler, M., and Schomburg, D. (eds.): Computer Science and Biology, Proceedings of the German Conference on Bioinformatics, GCB '96, (University of Leipzig, Leipzig, 1996.
15. Science, July Issue, 1996.
16. Boar, B.H.: The Art of Strategic Planning for Information Technology, (John Wiley & Sons, Inc., New York, 1993).
17. Parker, C., and Case, T.: Management Information Systems: Strategy and Action, (McGraw-Hill, New York, 1993).
18. Burkholz, H.: The FDA Follies, (Basic Books, New York, 1994).
19. Avishai, B.: Social Compact, Version 2.0. The American Prospect, July (1996), 28--34.

11. Disclaimer