December 2003

Implications of Information Technologies in Geo- and Bio-Sciences: A Literature Review

David W. Cheney
Silver Spring, MD

This literature review was prepared by David W. Cheney as an independent consultant.

The review was funded by National Science Foundation PO# D000141.

This report and a related issue brief are available at:
http://www.sri.com/policy/csted/reports/sandt/it

Questions or comments may be addressed to:
David W. Cheney at Dcheney@compuserve.com or cheney@wdc.sri.com

Acknowledgments

The author would like to thank the following individuals for their contributions to the report. Lee Regan of the U.S. Geological Survey, provided expert assistance in identifying and locating valuable resource materials, and provided helpful comments on the draft report. Alexandra S. Fairfield, formerly of the National Institutes of Health reviewed the draft report and provided assistance throughout the process. Daniel Rothbart of George Mason University, John Walsh of the University of Chicago, and Rachelle Hollander of the National Science Foundation reviewed all or parts of the draft report. Eileen Collins, formerly of the National Science Foundation, provided support and guidance throughout the process of preparing the report. Mary Frase of the National Science Foundation monitored the project in its final stages. The author appreciates the help of all of these individuals, but all errors and omissions remain the responsibility of the author.

About the Author

David W. Cheney is an associate director of the science and technology policy program at SRI International. He was vice president at Internet Policy Institute and was a senior executive in the U.S. Department of Energy, where he served as the director of the Secretary of Energy Advisory Board and as an advisor to the Deputy Secretary of Energy on industrial partnerships and the national laboratories. He worked for the private sector Council on Competitiveness, the Optoelectronics Industry Development Association, the Competitiveness Policy Council and the Congressional Research Service. He was a visiting researcher at Saitama University and has degrees from MIT and Brown University.

Contents

Acknowledgments ii
About the Author ii
Contents iii
Executive Summary v
  IT and Geoscience vi
  IT and Bioscience vii
  Comparing the Role of IT in Geoscience and Bioscience vii
  Policy Issues ix
  Gaps in the Literature ix
Introduction 1
  Purpose 1
  Scope and Methods 2
IT in Geoscience 4
  Evolution of IT in Geoscience 4
  Applications 5
    IT-Aided Instrumentation 6
    Modeling, Simulation, and Visualization 7
    Imaging/Tomography 10
    Geoinformatics 11
    Internet, Electronic Communication and Collaborative Tools 17
  Other Implications of IT for Geoscience 18
    Workforce issues 18
    Productivity 19
    Concept of Earth System Science 19
IT in Biosciences 19
  Evolution of IT in Biosciences 19
  IT and Molecular Biology 21
    Gene Identification and Annotation 22
    Comparative Genomics 23

- Protein Structure 24
- Gene Expression 26
- Applications 27
- IT and Biological Systems 29
  - Cellular and Subcellular Levels 30
  - Tissues/Organs 31
  - Organisms 32
  - Populations 33
- Cross Cutting Topics 34
  - Imaging and Instruments 34
  - Artificial Intelligence 35
  - Collaboration and Communication 35
  - Artificial Life 36
  - Workforce Issues 37
- Conclusion 37
- References 42
  - Geoscience References 42
  - Bioscience References 53

Executive Summary

This report reviews literature on the role of information technology in two fields of science: geoscience and bioscience. It was prepared as part of a National Science Foundation program to gather information about data and research on the social and economic implications of information technology (IT)[1] and to make it available for use by researchers, analysts, and the public.

Many information technologies, from the earliest computers to email and the Internet, have been developed for scientific applications. Understanding the role of IT for science is important because of the central role of science in today's information society. Advances in science affect economic performance and the achievement of societal goals, from health to national security. As a result, it is important for policy makers to understand how IT affects the quality and productivity of science.

Sociologists of science have long noted the importance of scientific equipment and instruments in leading to discoveries in science. IT may be viewed as an especially important kind of scientific equipment: it is used in virtually every field of science, and it has many diverse applications in data collection, analysis, and management; modeling and simulation; and communication.

There are several complementary approaches to examining the implications of IT for science. One approach is to examine the role of specific information technologies, such as email or the World Wide Web. Another approach is to examine the role of IT in specific aspects of the scientific process, such as scholarly communication or scientific collaboration. This report takes a third approach of examining the use of IT as a whole in specific fields of science. The advantage of this approach is that it looks at how multiple applications of IT simultaneously affect the conduct of science, and it also looks at how the implications of IT differ across fields of science. This approach also takes a science-centered, rather than IT-centered, view of the role of IT in science.

The geosciences and biosciences were chosen for this study because they differ with respect to their history and research methods in ways that one might expect would affect their use of IT. The biosciences have been largely an experimental science, whereas the geosciences have been largely observational -- based on observing what occurs in nature -- because of the impracticality of doing experiments on geologic spatial or temporal scales. The geosciences were early users of IT for such purposes as statistical analyses and analyses of seismic data. The biosciences, by contrast, were until recently rather modest users of computing. In recent years, however, the use of IT in biology has expanded to the extent that the biosciences are now among the heaviest users of IT.

---

[1] For the purpose of this study, IT includes the combination of digital processing, data storage, and the transmission of digital signals through telecommunications networks, and the applications enabled by these technologies.

The geoscience review focused on the solid earth geosciences, e.g., not oceanographic or atmospheric science, while the bioscience review focused on biological research, e.g., not medical or agricultural applications. Both reviews focused on research, rather than related subjects like technology development or science education.

Based on extensive searches conducted through the main bibliographic databases in the fields (GeoRef, Biological Abstracts), the study found there is a vast literature on applications of IT in the geo- and bio- sciences but little analytical literature on the effects of IT on the productivity, quality, or practice of research in these fields. Although the study uses such analytical literature where possible, the study is primarily based on literature that reviews or provides an overview of applications of IT in geoscience and bioscience.

IT and Geoscience

IT has been used for a long time in the geosciences, for such purposes as analyzing seismic data, doing statistical analyses, and drawing maps. Today, IT is used in a wide variety of ways, such as:

- collecting data using networks of digital instruments, such as seismometers or space-based sensing;
- storage and managing data in databases, such as for remote sensing data, rock and fossil data, and earthquake data;
- mapping and geographic information systems (GIS), to integrate and display a wide variety of data;
- modeling of a very wide variety of earth processes, such as earthquake mechanisms; flow in the earth’s core, sedimentation, petroleum reservoirs, hydrology, geochemical cycles;
- visualization to display the results of modeling;
- accessing scientific information; and
- communicating and collaborating with other scientists.

IT appears to be contributing to several changes in the geosciences. First, IT has played a part in the emergence of "earth system science" as an integrative view of the geological, oceanographic, atmospheric, and environmental sciences. This concept was largely driven by the availability of space-based remote sensing, but also relies on large shared databases and modeling of large-scale earth processes. Second, modeling has led a shift from geology being an observational and descriptive science to one that is more predictive. Finally, IT has also led to a change in the representation of geologic information, from static paper maps to interactive digital databases and geographic information systems.

IT and Bioscience

Biology is at its core an informational science based on the information embedded in the genetic code. At the molecular level, a key challenge in biology is to understand how instructions encoded in genomes lead to the structure and function of the proteins, and the regulation and expression of genes. IT is essential to store, manage, and decipher the mass of information produced in work in these areas. At levels of organization above the molecular level, computer models are critical tools to handle the complexity of relationships in biology.

Today IT plays a number of different important roles in bioscience:

- IT-aided instruments, such as gene sequencers, DNA microarrays, and microscopes, are used for data collection.
- Large shared, Internet-connected databases, such as gene and protein data banks, allow many researchers to obtain and contribute data to large problems.
- Data analysis methods are used in assembling and searching gene sequences.
- Imaging and visualization, such as Magnetic Resonance Imaging (MRI) and tomography are important in several areas of bioscience.
- Modeling and simulation are used to model protein folding and to help understand complex biological systems such as cells, tissues, organisms, and populations.
- Electronic communication technologies are used to share scientific information and aid scientific collaboration.

Comparing the Role of IT in Geoscience and Bioscience

Table 1 lists examples of applications of IT in the two fields. As can be seen in table 1, each field uses similar basic information technologies, including IT-aided instrumentation, databases, modeling and simulation, electronic communication, data mining.

Table 1 Examples of IT Applications in Geoscience and Bioscience

Information Technology	Role	Geoscience Applications	Bioscience Applications
IT-aided instrumentation/data collection	Increases efficiency of data collection	Digital seismometers, Remote sensing	Gene sequencers, DNA microarrays
Data analysis	Interpretation of data	Analysis of seismic data; statistical analysis of fossils, rocks.	Gene sequencing algorithms, protein folding calculations
Shared databases	Makes data widely available	Seismic, remote sensing, digital map, and paleontological databases, others	Gene sequence, protein, and many other databases
Modeling/simulation	Aids understanding complex systems; Allows virtual experiments	Earth simulator, digital earth	Virtual cell, virtual heart, ecosystem models
Imaging/visualization	Makes complex data intelligible	3D reservoir models, earth simulators, earth tomography	Protein folding, medical imaging
Internet/WWW	Enhances communication & collaboration	Uses in all fields	Uses in all fields

There are substantial similarities in the effect of IT in each field. In both fields, IT has aided more efficient data collection, is a critical tool for handling complexity, has aided greater dissemination of information, and appears to have facilitated widespread collaboration around large projects, often centered on contributions to shared databases or models of complex systems.

At the same time, there are also substantial differences in the use of IT in the two fields. The role of IT in biology is unique because information contained in the genetic code underlies much of biology. IT in geology is unusual because much of geoscience data is spatial in nature and is enhanced by the use of geographic information systems. Modeling plays a different role in the two fields because in bioscience it is often possible to test models with experiments whereas the temporal and spatial scales involved in geology often make such testing impossible or impractical.

IT appears to have influenced scientific processes in both fields. The availability of large datasets (e.g., gene sequence, protein structure, and remote sensing data) is enabling scientists to conduct research using data developed by others, rather than generating their own data. In addition, modeling and large shared databases facilitate work on large, systems-level complex problems, as opposed to studying small pieces of large problems. Systems biology and earth systems science rely on modeling of complex systems and large databases.

Policy Issues

There are several policy-related issues relating to the use of IT that were raised in the geoscience and bioscience literature. One issue concerns the use of models in policy decisions, such as decisions regarding the adequacy of nuclear waste sites, or the viability of ecosystems. These models cannot be fully validated by experiments, due to the temporal and spatial scale of the experiments that would be required, and there is debate over the appropriate use of these models. Other issues include intellectual property protection issues, such as the appropriate level of protection for databases, the development of standards necessary for interoperability of databases, the development of metadata to aid searching of databases, and issues related to having adequate people with appropriate training for working in IT in science.

Gaps in the Literature

There have been few controlled evaluations or critical analyses of the role of IT in the geosciences and biosciences. While there are studies of the use of specific aspects of IT (such as electronic scholarly communication and IT-enabled collaboration) in science, there is little literature that analyzes the role of IT in changing research methods, research questions and directions; changing the productivity and quality of science; or changing the organization of science. This suggests a large potential agenda for future work.

Introduction

Purpose

This literature review examines how information technology (IT) is being used in the biosciences and geosciences. For the purpose of this study, IT was defined to include digital processing, data storage, and the transmission of digital signals through telecommunications networks, and the applications enabled by these technologies. The geosciences and biosciences were chosen for this study because they differ with respect to their history and research methods in ways that one might expect would affect their use of IT. The biosciences have been largely an experimental science, whereas the geosciences have been largely observational, because of the impracticality of doing experiments on geologic spatial or temporal scales. The geosciences were early users of IT for such purposes as statistical analyses and analyses of seismic data. The biosciences, by contrast, were, until recently, rather modest users of computing, but in the past two decades the use of IT in biology has expanded greatly. By comparing the use of IT in such diverse fields, it is possible to generalize to some degree about the role of IT in science.

The role of IT in science is of interest for several reasons. First, it is important to understand developments that may change the rate, direction, and nature of science. It has often been noted that scientific equipment or instruments play an important role in determining the goals, research methods, and productivity of science. IT may be viewed as an especially important kind of scientific tool because it is used in virtually every field of science, and because of the diversity of its uses in such areas as data collection, analysis, and management; modeling and simulation; and in communication. If IT is influencing the rate or direction or quality of science, it is important for policy makers and educators, as well as scientists themselves, to understand these changes.

Second, studying the role of IT in science may provide insight into the broader implications of IT for society. There have been many analyses of the effects of IT in business, education, and the home, but there has been relatively little analysis of IT and science. Analyses of the role of IT in business and education have shown that benefits from IT were often much less than initially predicted, at least in the short-run. Achieving benefits, such as improved productivity or improved educational outcomes, often required additional changes, such as in changes in organization or investments in training. Do the same observations hold true for science? If so, it would support theories about how IT affects various application domains. If, on the other hand, the implications of IT for science are markedly different from the implications for education or business, an analysis of the differences could help expand or enrich the theories related to the social and economic implications of IT.

Third, knowledge about role of information technology in science is relevant to several public policy issues. Among these are:

- R&D funding issues, such as how much Federal support for science should be devoted to IT, versus traditional science and other scientific infrastructure;

- workforce issues, such as how IT affect the skills and specialties needed to conduct science; and

- the appropriate use of predictions based on modeling and simulation in public policy decisions.

This study found that although there is a vast literature on applications of IT in geoscience and bioscience, there is very little analytical literature that addresses the questions discussed above. There are few assessments of how IT has changed the productivity, quality, or practice of research in these fields. As a result, this study is largely based on a review of literature that describes the uses of IT in geoscience and bioscience. It sheds light on a somewhat narrower set of questions:

- How has IT been used in the two fields, and which applications (e.g., modeling and simulation, shared databases, collaborative tools, remote instrumentation) have been most widely used?

- How is the use of IT similar and different in two diverse fields of science?

- What are some of the major effects and implications, such as the development of new subfields or changes in research methods?

- What are some of the major issues in the uses of IT?

Scope and Methods

The scope of this review was defined to be literature limited to:

- English language literature.

- Literature published between 1992 and 2002. A few earlier publications were reviewed for historical perspective. In cases where many publications covered the same topic, the review focuses only on the most recent publications.

- Articles focusing on IT used in the conduct of science, rather than related subjects such as science education.

The selection criteria for the bioscience and geosciences reviews were slightly different, reflecting differences in the size and nature of the fields.[2]

---

[2] For example, in 1999, U.S. R&D expenditures at academic institutions for life sciences was $15.9 billion, compared to $1.7 billion environmental science (including earth, atmospheric, and ocean sciences). National Science Board. "Appendix table 5-7, Total, Federal, and non-Federal R&D

For the geosciences, bibliographic searches were conducted on GeoRef (Cambridge Scientific Abstracts), the main geoscience bibliographic database. Search terms included "computer simulation*", "computer model*," "visualization," "Internet", "data management" (where "*" indicates a wild card symbol, capturing any combination of letters). These yielded a large number of hits. For example "data management" on GeoRef yields over 1000 hits and "Internet" yielded over 900. The citations and abstracts were reviewed to select out a representative number of review articles. Most materials were accessed using the U.S. Geological Survey library in Reston, VA.

The geoscience review was limited to focus on the hard earth sciences, such as geology and geophysics, rather than oceanographic, atmospheric, and space science.[3] The review included:

- articles in peer reviewed journals;
- articles in professional journals (such as *Geobyte* for petroleum geologists) that are not peer reviewed but describe techniques used by practicing scientists;
- professional papers and government publications (such as U.S. Geological Survey Open File reports); and
- abstracts in proceedings of professional society meetings.[4]

No attempt was made to review doctoral dissertations. Several books that were identified in the course of the study were reviewed, but no attempt was made to comprehensively cover books.

For the biosciences review, searches were made primarily using Ovid Biological Abstracts, which covers articles from over 7,000 journals, as well as original research reports and reviews, letters, and technical data reports.[5] Searches on subjects such as computers or computational biology resulted in a huge number of hits. For example the “Computational Biology” subject heading resulted in 106,352 hits. "Computer%" (where % is a wild card) resulted in a similarly large number of hits. Other search terms used included "modeling" and "bioinformatics". Due to the large number of hits, searches were then limited to (1) review articles; and (2) articles in English. This reduced the

---

expenditures at academic institutions, by field and source of funds: 1999". *Science & Engineering Indicators – 2002*. Arlington, VA.

[3] While this scope for “geoscience” was consistent with that used by some studies of IT in geoscience (e.g Loudon, 2000), it also excludes some areas in which IT is used heavily, such as climate and oceanic modeling.

[4] Many discussions of the use of IT in the geosciences were presented as talks in professional society meetings that were recorded only as abstracts in the meetings proceedings.

[5] Biological Abstracts was used rather than more biomedical oriented databases (such as PubMed) because the reviews focus is on bioscience rather than biomedicine.

number of hits to the low thousands. The majority of these articles were narrower in focus than desired, but articles that provided overviews of significant applications of IT in biology were selected. Where there was duplication in the coverage, more recent articles were selected.

The bioscience review was limited to the effects of IT on biological research, rather than the application of IT in such biology-related applied fields as public health or agriculture.[6] For example, the review covers IT in medical research but not the application of IT to the practice of medicine. Also, the review does not cover the effects of biology on computing, such as biochips or bio-inspired computing. In addition, it focuses on IT in computing and communications, not on the use of microprocessors in lab equipment.

IT in Geoscience

Evolution of IT in Geoscience

Computers have been used extensively in the geosciences since the 1950s (Krumbein 1962, Wagner 1992, Davis and Herzfeld 1993). By the early 1960s, it was recognized that computers had many applications in geology, including automatic data acquisition, data storage and retrieval, summarizing large masses of data, and limited analysis by more complicated computational methods (Krumbein 1962). Early computers were also used for presenting data, including the plotting of scatter diagrams, regression lines, contour maps, and other visual output. The use of computers to produce smooth approximations of three-dimensional surfaces was especially valuable in the petroleum industry, in order to represent subsurface features based on well data (Davis and Sampson 1992, Wagner 1992.).

From the early days, it was recognized that geology was frequently an observational science, i.e, observing what occurs in nature and trying to make sense of it, and only sometimes an experimental science (Krumbein 1962). As an observational science, geology was often qualitative, and was characterized by many complex variables and inadequate data. As a consequence, statistical methods were often the appropriate quantitative techniques. Computers were used for data management and analysis, including classification of fossils, rocks, and landforms; analysis of relationships between dependent and independent variables; analysis of area variations (map analysis); and associations among variables (analyses of statistical correlations) (Krumbein 1962). Many kinds of computer statistical analysis (factor analysis, regression, etc.) were especially useful.

The International Association for Mathematical Geology (IAMG) was founded in 1968. It expanded beyond its original focus of applying statistical models to geological

---

[6] This criteria was applied more restrictively in the biosciences than geosciences. The geoscience review discusses some uses of IT in application of geoscience in the petroleum industry, for example.

data, and now focuses to a large degree on information technology applications in geology. In 1975, the IAMG launched the journal *Computers & Geoscience*, which has become the main journal at the intersection of IT and geoscience.

Statistical methods are still core computer techniques in geology, and statistical computer applications, such as regressions and factor analysis, are used in a wide range of applications, from understanding climate change to conducting risk analysis of petroleum exploration. Over time, these applications have migrated from mainframes to PCs (Davis and Herzfeld 1993). More recently, as IT has improved, many other applications have come into widespread use. Among these are modeling and simulation, visualization, and large shared databases. With the advent of the Internet, large data sets are shared among scientists (and the public), and research results and publications are made available electronically.

The petroleum industry has been an early and intensive user of information technology. The industry used computing for the processing and management of seismic survey data and other exploration data (Dawson and Lim, 1993). Prior to 1980, information technology was used mostly in areas that required intensive numeric calculations, such as seismic processing and mapping, and was done by technical specialists. In the 1980s, the emergence of minicomputers, online editors, and graphics devices led to interactive applications. It then became possible to manipulate geologic interpretations on screen (Dawson and Lim 1993). By the mid-1980s, commercial database management systems came into widespread use in exploration companies. By the late 1980s, most exploration companies had invested heavily in information technology, making use of central computing (and sometimes supercomputing) as well as desktop systems. Many systems were not compatible, and the industry has moved toward common software environments based on standards.

Applications

Today, IT is used in a wide variety of applications in the geosciences. Table 2 provides an overview of these applications.

Table 2

Technology	**Role**	**Geo-Application**
IT-aided instrumentation and data networks	Connects instruments	Seismometers, space-based sensors
Modeling/Simulation	Tool for understanding complex systems; Allows virtual experiments where real experiments are impossible or impractical	Climate, earth interior, tectonics, sedimentation evolution
Visualization	Makes complex data intelligible	Subsurface formation, reservoirs, climate models
Tomography	Imaging of interior of objects	Earth interior, fossils
Shared Databases	Makes data widely available	Maps, remote sensing data, oil and gas data, paleontology databases
Mapping/GIS	Integration of geospatial data	Rock formations Resources, hazards, land use
Collaborative tools	Enhances communication between researchers	Distributed research projects
World Wide Web	Makes data, models, preprints, educational materials easily available	Public dissemination of information, education
E-journals	Speeds and facilitates access to literature	All fields

The following sections describe these applications in more detail, and provide some examples.

IT-Aided Instrumentation

Information technology is contributing to better collection of data in many areas.[7] One example is seismology. Digital seismographs replaced analog seismographs beginning in the mid-1970s (Romanowicz and Giardini 2001). As a result, more complete information, including full seismic wave forms, are routinely readily exchanged through a variety of media, including satellites, digital phone links, and especially the Internet. The more extensive and complete information has enabled tomographic investigations of the Earth's interior structure. In addition, earthquake surveillance relies

---

[7] The focus of this analysis is not on every application of microprocessors in scientific instruments, but instead on larger computation and networking.

on real-time data acquisition from national and local networks. Rapid access to high-quality seismic waveforms has also allowed detailed investigations of large disastrous earthquakes to be conducted much more quickly after the event.

The higher quality information from seismographs around the world has also supported detecting underground nuclear explosions, which is essential for monitoring nuclear test ban treaties (National Research Council 1997). Detecting small nuclear tests and discriminating between such tests and earthquakes and mining related events is a challenging problem that relies on advanced seismic networks (Walter and Hartse 2002). A central part of the monitoring regime for the Comprehensive Nuclear-Test-Ban Treaty (CTBT), as adopted by the United Nations General Assembly, is the global network of digital seismographs.

Another area where IT-enabled instrumentation has had a large effect on the earth sciences is space-based remote sensing. Space borne remote sensing data platforms, such as the Space Shuttle and satellites, provide data for a wide range of geologic applications (Morgan et. al. 1997). There has been a great increase in the capabilities of these systems in the last few years (Kafatos et. al. 1999). Global earth observing missions and operational satellites produce large volumes of public domain data. The Internet and the World Wide Web allow these data to be accessed by a variety of scientists, applications experts and the general public.

IT has also enabled improved data collection in other ways. For example, Internet surveys have been used to collect earthquake intensity data from the general population[8] (Cajka and Halchuk 1998).

Modeling, Simulation, and Visualization

Long before the advent of computers, geologists used physical models to explain or illustrate geoscientific phenomenon such as stream flow (Brice 2000). Conceptual models, such as plate tectonics, were formulated without the use of computers. The use of computers, however, has greatly expanded the use of geologic modeling, and computer modeling is now used to explore a large number of geological processes. Modeling is especially useful in areas where experiments are difficult or impossible to conduct. This applies to many geological phenomena that take place on spatial or temporal scales for which experimentation is impractical or impossible.

The geosciences have often been among the intensive users of supercomputers. Several geoscience topics, including mantle convection, oil recovery, and ocean turbulent flow were among the early “Grand Challenges" for supercomputers in the U.S. High Performance Computing and Communications Initiative (Grossman 1992). In 2002, the world’s fastest computer, the Japanese 40 teraflop (40 trillion floating point operations

---

[8] *Intensity* measures the effects of the earthquake shaking at a given site and thus requires observations of the earthquake effects at specific sites, as opposed to *magnitude*, which measures the energy of the earthquake, and is determined from seismometers.)

per second) “Earth Simulator,” was devoted to earth science simulation s (Normile, 2002).

In the geoscience literature, the term "model" can refer either to static models that mainly provide a visual representation of data or to dynamic models that illustrate or predict the action of complex processes. The static models typically provide a three dimensional view of rock formations and allow a vast amount of geologic or seismic data to be shown. Computers are used to process the large amounts of data and generate the images rock of formations that represent the spatial relationships. Once a model is constructed, it is available to generate a variety of maps and cross sections. These have been used extensively in the petroleum industry for mapping seismic and well data (Jones and Leonard 1990). The models and graphics allow researchers to get a better feel for the geometry of underground formations, and allow testing of alternative interpretations of the data set (Flynn 1990). Models have increased in complexity as computer power has increased. Many 3-D modeling techniques were originally developed for mechanical engineering and other man made objects. These techniques had to be adapted to geoscience, where there is limited sampling and the objects being modeled are irregular and complex (Fried and Leonard 1990).

Dynamic models show geologic processes over time. In many cases, quantitative computer models are the only way to simulate of the interaction of many process and factors over time (Vining 1998). Table 3 shows examples of many of the applications of dynamic models in the geosciences.

Table 3

Application of Model	**Reference**
Global climate	Smith et. al 2002; Normile 2002
Ice streams and glaciers of the Antarctic ice sheet	Takeda, Cox, and Payne 2002
Interior of the earth (core, mantle, etc.)	Normile 2002
Mineral properties	Refern 1995
Formation of crystal structure in igneous rocks	Amenta 2001
Simulations of the distribution of fossils	Holland and Patzkowsky 1999
Oil migration through rock	Souto Filho 1998
Development of ancient rock formations	Mancini, Chen, and Benson 1998
Creation of mineral deposits and tectonics in the Pacific	Scotese et. al 2001
Petroleum reservoir formation and evolution	Kacewicz 1997
Relationship between fault movement and sedimentation	Smith-Rouch et.al 1997
Geochemical reservoirs and fluxes in the earth	Staudigel et. al 1998
Evolution of river systems and watershed interactions	Vining 1998

Dynamic models include both inverse models, for determining processes from observed data, as well as forward modeling, for studying the interaction of processes to produce a response (Rankey and Watney 1994). They also include both deterministic simulations, in which a process leads directly to a known response, and stochastic simulations, which contain an element of chance and may or may not produce the same result each time (Rankey and Watney 1994).

Closely related to modeling is visualization. Visualization in this context is the use of information technology to present complex data in a visual form, frequently in three dimensions and often over time. In most cases, visualization presents the results of models or simulations in a form that lets the user see patterns in the data. Much of the human brain is dedicated to visual processing of information, and people can usually detect patterns more readily when they are displayed in visual form (Zeitlin 1992). Visualization technology, like modeling, has been used in a wide number of applications. These include:

- the visualization of subsurface geology from wireline logs (Collins 1998);
- visualization of mineral structures (Lennie 2000);
- subsurface fluid migration (Birken and Versteeg 2000);
- evolution of petroleum traps and reservoirs (Kacewicz 1997);
- reservoir depletion (Costello1998);
- global warming (Gordin and Edelson 1997); and
-
- oil exploration and production (Zeitlin 1992).

There is a discussion in the geoscience literature about the appropriate uses and benefits of models (Rankey and Watney 1994, Smith et. al. 2002, Staudigel et. al. 1998, Scotese 2001, Oreskes 2000, Oreskes et. al. 1994). Several key points emerge from that discussion:

- Models can provide new, sometimes non-intuitive, perspectives on the data, and assist in integrating available information. They can lead to major new interpretations of systems, and can challenge and refine understanding of geologic systems.

- Models encourage accuracy and precision in data collection and interpretation by establishing formal quantified assertions in logical terms. Discussions around models can focus the testing of models, highlight gaps in knowledge, and trigger more focused work and data collection. Models often identify the need for new data. In climate models, for example, errors evolve as a result of incomplete physical understandings and limited knowledge of past (or future) climate forcing. Modeling highlights the areas where more data and knowledge is needed.

- Models can be used for sensitivity analysis, to show where weaknesses in data or theory are most serious.

- Models are tools for building consensus. They serve as a focus for communication between dispersed parts of the technical community. For example, the geochemical reference model serves as a focal point for discussions in the international community (Staudigel et. al. 1998).

- By making knowledge explicit, models can make the field more transparent to the non-specialist. Models can stimulate interdisciplinary cooperation and promote scientific and technical exchanges (Scotese 2001). Visualization technologies can also aid in communication by displaying information in form that is understandable across disciplines.

Oreskes (2000) cautions against the improper use of models for prediction, especially in areas that have important public consequences, such as in predicting climate change. Geoscience models generally cannot be verified in the sense of being shown to predict reality accurately. No model output ever matches the world exactly, in part because the input data is never complete. The real world (especially in geosciences) is heterogeneous at different scales. Input data for models generally provides a single average number for a cell that is in fact heterogeneous.

In addition, although the ability to represent complex systems is a key strength of models, it also poses a challenge. The more complex a model, the more parameters can be adjusted to calibrate the model so that its predictions match observations. More than one complex model can be calibrated to match data, however, and models that do not accurately match reality can nevertheless be calibrated to data (Oreskes, 2000). If models are not based on accurate theoretical foundations, they are unlikely to be accurate outside of the range for which they have been calibrated. All of these aspects of models mean that they should be used with great caution in making predictions that are the basis for public policy.

Imaging/Tomography

Imaging techniques that rely heavily on information technology have been used in a number of different applications in the geosciences. Seismic reflections have long been used to assess underground formations, and, more recently, tomography has been used to assess the inner structure of the whole earth. A variety of technologies can potentially be used for rapid, relatively inexpensive noninvasive characterization of the Earth's subsurface (National Research Council 2000). Non-invasive techniques include seismic and electromagnetic surveys and remote sensing from aircraft and satellite. The capacity to characterize the subsurface environment is important to supporting buildings and other structures, finding water, and managing waste. IT is critical to these noninvasive techniques, for data acquisition and data processing as well as for modeling and visualization. Borgman (1998) suggests that information technology has contributed to the increasing transparency of the earth. For most of human history, the Earth has been solid and impenetrable. More recently, the solid Earth and the oceans, have become more accessible and intelligible, making the Earth seem more transparent.

Another geoscience application of imaging is the use of medical CT scan technology to determine the internal morphology of fossils and to aid in reconstruction and modeling (Gould et. al 1996). Anderson et. al. (2001) describe the use of rapid protoyping technology, using a variety of prototyping techniques and materials, to create models of fossils.[9] In this process, fossil materials are scanned and three-dimensional virtual models are generated. Rapid prototyping technology is used to create missing pieces (where there is bilateral symmetry) or to create miniatures. Rapid prototyping has been combined with C-T scanning to create models of fossil braincases.

Geoinformatics

Much geologic work -- such as remote sensing, seismology, and oil exploration -- produces huge amounts of data. The term “geoinformatics” has come into usage to describe the management and storage of geological information. IT provides the capability to store massive volumes of such data and deliver it through a variety of networks. A wide variety of databases and information management systems have become vital to the geosciences. These include such diverse systems as databases of fossils and rock types, databases of satellite remote sensing data, petroleum exploration databases, and databases that contain the information traditionally represented in geological maps.

A great deal of geological information is spatial. In some cases, such as information about size and location of a geologic structure, the key information is inherently spatial. In other cases, such as a database of fossil types, much of the information may be aspatial, but there is spatial information is associated with each of the records, (e.g. describing where a particular fossil was found). Much geological information has traditionally been both stored and represented in paper maps. A major influence of information technology has been the movement of geological information from paper maps to databases that can present the data in maps as well as in other forms.

There is movement towards integration of the these databases towards a "digital earth," through which a user can find a very wide variety of geological, paleontological, mineralogical, seismological, hydrological, meteorological, and other information about a location or region through a single database. Such integrated databases have become important both for scientific research and for applications of geoscience information in economic geology, land use planning, natural hazard mitigation, and other areas. The next sections describe topical databases, geographical information systems and mapping, and some issues related to geoinformatics.

---

[9] Rapid prototyping technology refers to using computer aided machines to create three dimensional models from computer-based designs. It was developed for use in manufacturing to develop prototypes of new designs.

Topical Geoscience Databases

There are a wide variety of geoscience databases, including databases of fossil and rock types. Kaiser (2000) notes that although fossil database have been slower to develop than, for example, genomic databases, there are now numerous fossil databases. Examples of paleontological databases include:

- The U.S. Geological Survey National Paleontological Database (Wardlaw et. al. 2001).

- The Plant Fossil Record database (Lhotak and Boulter 1995, Boulter 1999), which is being built to include taxonomic, morphological, geological, geographical, and bibliographic information of all authoritative plant fossil citations throughout the world. The database facilitates comparison of different hierarchical taxonomic models and testing of theories of evolution and biogeographical migration. http://bs.uel.ac.uk.ibs/

- The global pollen database (GPD), which contains Quaternary[10] pollen data from around the world. The North American pollen database, European pollen database, and Latin American pollen database have contributed to the GPD. The GPD contains both archival data, which is the original data, and research data, which involves manipulation of the archival data (Grimm and Keltner, 1998).

- The Catalogue of Palaeontological Types in Austrian Collections (http://www.oeaw.ac.at/oetyp/palint.htm) database provides quick access to basic information about the scientifically most important paleontological specimens kept in Austrian institutions.

Examples of rock databases include:

- The Igneous Data Base and a Sedimentary Data Base managed under the Auspices of the International Union of Geological Sciences. The igneous database contains about 20,000 entries of geochemical and mineralogical compositional data of volcanic and plutonic rocks. The Sedimentary Data Base contains about 5,000 entries of chemical, mineralogical and granulometric compositional data of various sedimentary rocks, together with sedimentological and geotechnical descriptions (Brandle et. al. 1997). Both are available through the Internet. (http://www.ige.csic.es/sdbp/sdbp.htm)

---

[10] The Quaternary is a subdivision of geological time (the Quaternary Period) which covers about the last two million years up to the present day.

- The Stanford Rock Physics laboratory has collected data on the acoustic properties of rocks over many years (Nolen-Hoeksema and Hart 1991). The data has been put into a relational database that allows flexible searching.

- RidgePetDB and GEOROC are comprehensive, searchable on-line petrologic and chemical databases for ocean floor and oceanic island rocks. They include a broad spectrum of supplementary information (metadata) that describes the quality of the analytical data, sample characteristics, and sampling process that can be used to evaluate, filter, and sort the chemical data (Lehnert et. al 2000).

- The Ocean Drilling Program operates the JANUS database. The database includes paleontological, lithostratigraphic, chemical, physical, sedimentological, and geophysical data for ocean sediments and hard rocks (http://www-odp.tamu.edu/database/).

There are a variety of websites that provide more extensive lists of Internet accessible data resources, including GeoGuide (http://www.geo-guide.de/info/index.html) and EarthRef (http://earthref.org/erlinks/main.htm).

Geologic information has been especially important in the oil industry (Arthur 1996). Petroleum exploration produces vast volumes of seismic, well, and other data in a variety of media and formats (Kingston and Simson 1997) and there are a wide variety of data bases related to the oil industry. States and oil and gas producing companies have used electronic databases for decades. Early ones were non-relational and difficult to use. More recently, PC-based, fully-relational, normalized, and comprehensive have been developed. Some examples of petroleum-related databases include:

- The Kansas Geological Survey's Digital Petroleum Atlas (DPA) project (Adkins-Heljeson and Carr 2001). The DPA consists of reservoir information from the individual well bore to regional studies. It presents information such as structure maps, cross-sections, and petrophysical and core analyses in web pages and relational databases. Users can navigate through the pages from the regional level to the borehole.

- The American Geological Institute (1997) has established the National Geoscience Data Repository System (http://www.agiweb.org/NGDRS) to preserve geoscience data made available due to the downsizing of U.S. petroleum and mining companies.

Other important parts of the geologic data landscape are the World Data Centers (Stoss 1998). World Data Centers were established in the United States and elsewhere around that world as an outcome of the International Geophysical Year in1957-1958. They have been the formal mechanism for the international exchange of information in disciplines related to the earth, its environment, and the sun. Today there are 46 individual data centers within five major World Data Center complexes. Data is available through in person visits, mail request, or through online communication.

Thirteen discipline-based World Data Centers and a Coordination Office are located in the United States. The National Oceanic and Atmospheric Administration operates several of these, including the National Geophysical Data Center (http://www.ngdc.noaa.gov/ngdc.html). These distribute data and information about solid earth geophysics, glaciology, marine geology, solar and terrestrial physics, and paleoclimatology. Over the last 25 years data distribution has evolved from printed data catalogs and mailing of tapes, through online catalogs and FTP distribution, to Internet- and Web-based access and search tools (Habermann and Anderson 2001).

Geographic Information Systems and Mapping

Much geoscience information is spatial in nature and maps have been an important way of representing geologic knowledge. Geographic information systems (GIS) and digital mapping have become important tools for managing and displaying spatial geoscience data (Heron 1994). Maps are also important media for communicating information about geologic and resources to user groups. Geologic maps are used in risk assessment, hazard identification, and other practical social planning functions that link geology to other geographically distributed data.

Geologic mapping has evolved from the early hand-inked paper maps through the era of photogrammetry and remote sensing, to the digital age (Kramer 1997). Traditional geological maps convey the mapmaker's understanding of the geology of an area and also serve as the *de facto* repository for much of the recorded geology of the area (Giles and Bain 1995). Traditional geologic maps, however, are difficult and expensive to update. Many areas were mapped a long time ago, and advances in geologic knowledge have occurred since many areas were mapped. In addition, additional information, such as information on mineral or water resources, or on the geology and geochemistry at each site, becomes available. As a result, maps need to be revised from time to time to reflect new data or interpretations of the data.

Geographic information systems (GIS) overcome many of the disadvantages of paper maps. GIS systems can be updated frequently and can be used to store, integrate, interrogate, and analyze different types of data from different sources. It is now widely recognized that the principal product of geologic mapping should be a database, from which both analyses and traditional maps can be derived (Black and Walker 2001, Giles and Bain 1995). Centrally managed geoscientific databases (Bandy and Wallace 2001) can:

- be easily maintained;
- be queried to find information;
- provide improved access, including use by multiple users at the same time;
- provide information in a single source; and
- integrate spatial and aspatial data.

Work is ongoing in many places around the world to develop digital, database-centered, geological maps. In the United States, an important effort is the development of the National Geologic Map Data Base (http://ngmdb.usgs.gov) (Soller 1997, Soller et. al. 2001), which is a collaborative effort between the USGS and the Association of American State Geologists (AASG). The Database was mandated by the Geologic Mapping Act of 1992 to serve as an archive of spatially-referenced geologic, geophysical, geochemical, geochronological, and paleontological information. The database is being developed in several phases that include development of:

- a searchable catalog of all paper or digital geoscience maps in the United States;
- standards and guidelines for the management of digital map information in the database; and
- standardized regional and national-scale geologic map coverages.

The map catalog is supported by two databases: GEOLEX, a searchable geologic names lexicon, and Geologic Mapping in Progress, which provides information on current ongoing mapping projects.

When the project began in 1993, computer-based mapping was not sufficiently advanced to allow the development of a national online map database, and the World Wide Web was not yet widely available to the general public. Now, however, advances in digital geologic map data coverage, standard data formats, data models, digital mapping practices, field data capture techniques, and Internet delivery of spatial information have allowed the building of a prototype online National Geologic Map Database (Soller et. al. 2001).[11] Another related resource is the USGS Geoscience Data Catalog, which links to metadata records and/or actual digital products that can be accessed online (http://geo-nsdi.er.usgs.gov/).

Complementary work is proceeding in many states and countries. Most European geological surveys are producing digital maps, but efforts are limited by a lack of standards (Jackson and Asch, 2002; Anderson et. al. 1999; Chirico 1999; Stanford and Freed 1997; Pristas and Herman 1997; Richard 2000; .Mitchell, Fritz and Waldkirch 1999). Efforts are underway to achieve global coverage, such as through the Global Mapping Project (Une 2001), the Global Spatial Data Infrastructure, Digital Earth Project and the UN Geographic Information Working Group.

Ideally, the digital information for geologic maps should be created in the field, as the geologic observations and interpretations are being recorded. Field geologists now have access to a variety of tools for digital geologic mapping, including lightweight rugged field computers, global positioning systems (GPS), laser range-finders, and digital photography linked by data acquisition software (Kramer 1997). Field geologists can now

---

[11] For information on the digital map project, see http://ncgmp.usgs.gov/ngmdbproject/.

create digital maps and cross sections in the field while making observations. Many field geologists, however, do not yet record their observations in digital form, due to cost and lack of familiarity with the wide variety of hardware and software tools (Black and Walker 2001).

A goal is for geoscience information systems is to move toward the concept of the "digital earth" in which geoscience data is available universally in digital form and in a form that can easily be searched and integrated (Zhao et. al. 1999). A regional implementation of this concept is Cornell University's Geoscience Information System Project, which has been applied to the Middle East (Seber et. al. 2000, 2001). The system aids the assimilation, and management of large and diverse data sets, such as geologic maps, faults, topography, and seismic events, volcanic activity, oil and gas fields, satellite imagery, and makes it available on the Web (http://atlas.geo.cornell.edu). Another implementation is the Global Earth Information System (Iwao 1998), which combines point features (wells, mines, epicenters) linear features (geologic boundaries, faults) area features (soil, rock types), and volumetric features (soil deposits, liquid volumes).

The U.S. National Research Council (1999) has recommended the development of an even broader concept -- a distributed geolibrary that would permit users to quickly and easily obtain all existing information available about a place that is relevant to a defined need. Part of this information would be geological, but it would also contain information relevant to a wide range of problems, including natural disasters, emergencies, community planning, and environmental quality. The NRC defines a geolibrary as a digital library filled with information associated with a distinct area or footprint on the Earth's surface, and for which the primary search mechanism is place.

Issues and Challenges in Geoinformatics

There are a number of challenges to building successful comprehensive geoscience databases. Integration of heterogeneous data sets is a formidable problem for geospatial data (Habermann and Anderson 2001). Research needs include problems of indexing, visualization, scaling, automated search and abstracting, and data integration or "conflation"[12] (U.S. National Research Council 1999). That ability is severely impeded today by differences in formats and standards, access mechanisms, and organizational structures. Data integration is even more challenging when it extends beyond geology and geophysics to engineering and economics (Davis 2000).

A key aspect of this is developing appropriate "metadata" or information about information, such as keywords, subject categories and other data attributes that aid in the searching of information. A key challenge is to determine appropriate ways to describe and document data with sufficient information so that users searching across distributed computer networks can discover the data, and both specialists and scientists in other fields can successfully use the data (Habermann and Anderson 2001). A related issue is

---

[12] GIS data conflation is the process of creating and updating a new "master" dataset from the best spatial and attribute qualities of two or more source datasets.

the development of standards for digital map production (Jackson and Asch 2002). A key issue is defining and agreeing on fundamental geological classification systems and procedures. Whereas paper geological map series will tolerate inconsistency between separate map sheets, the greater degree of integration in GIS and related digital systems requires a more stringent approach.

Intellectual property and database protection issues also affect the development of geoscience databases (Applegate 1998). Governments around the world have been working to broaden intellectual property protection for electronic databases. These new policies alter the balance between access and protection that has allowed the current scientific system to flourish. There is a long-standing debate over what form that protection should take.

Incorporating data from traditional types of documents and collections into digital information systems presents another challenge. Geoscience data sources that need to be upgraded and integrated into electronic form include: manuscript collections with unpublished or draft papers; unpublished technical information such as geochemical analyses or stratigraphic sections; and supporting documentation such as annotated aerial photographs, wire-line logs, and geophysical data (Brown and Love 1997). A further challenge is providing online information about three-dimensional artifacts, such as hand specimens, thin sections, and fossils, which are important in geoscience. These collections are often poorly maintained, managed and cataloged (Browne and Love 1997). These specimens are often separated from the manuscript collections to which they are logically connected, with the result that researchers may be aware of one portion of a collection and unaware of others.

Internet, Electronic Communication and Collaborative Tools

The Internet and World Wide Web are used extensively to share and disseminate geoscience information – both through posting information directly on the World Wide Web and through electronic journals. In the early days of the Web (around 1995) many articles appeared in the literature that pointed to new Internet resources for geoscience. For example, Last (1995) described Internet listservers of interest to paleolimnologists; Dunn and Feldman (1995) discussed bulletin board and web sites related to the history of geosciences; Ingram (1996) pointed to geoscience uses of the Web; Zwolinski (1996) pointed to Polish geoscience resources on the Web; Thoen (1995) described Internet resources related to GIS and mapping; and Ramshaw (1995) described general geoscience listservers and newsgroups.

Geoscience organizations are now using the Internet as the main means of getting information to their stakeholders. The U.S. Geological Survey (Wendt and Lanfear 1999) and the Kansas Geological Survey (Buchanan and Carr 1997, Carr et. al. 1996) each describe how the use the Internet to provide access to geologic, hydrologic, and geographic data, data compilations, and research and technical studies. Products published online are made available more quickly (as they are completed), at a lower cost, and to a wider user-audience. The Kansas Survey is designing research and technical products that go beyond traditional publications and take advantage of the Internet capabilities. The Internet speeds the dissemination of Survey's information and research results, aiding a variety of individual and public policy decisions.

Geoscience is also affected by general trends in academic publishing. Hallmark (1998) notes that current trends in geoscience journal publishing include higher prices for journal subscriptions, cancellation of print titles by libraries, increasing availability of electronic journals and indexes, and new communication patterns facilitated by the Internet. Derksen and Haner (1997) note that the great access to earth science information in libraries comes at a cost in workstations, high speed network connections, hardware and software security systems, printers and printer supplies, staffing, contracts for access to indexing products and electronic journals, and training expenses. These are substantial costs for libraries.

Some researchers have used collaborative tools to facilitate collaboration over the Internet. Burns (1997) describes Internet-based collaboration using a secure website in a study of natural rock fracture. The research is done by geoscientists at various research organizations in the United States and abroad and is supported by a consortium of energy companies. With the study's private website and associated network resources, all research study participants have access at their desktops to current project data, archived files and publications, and online forums. Burns notes that the use of Internet technologies has resulted in greater collaboration, closer relationships among project participants, faster progress, and reduced costs. Other examples of geoscience use of Internet-based collaboration include the use of a collaborative environment for Mars surface science studies (Gulick et. al. 2001) and the use and control of an X-ray diffractometer over the Internet (Argast and Corey 1998).

Models also serve as the basis for collaboration. For example, the Geochemical Earth Reference Model (see http://www-ep.es.llnl.gov/germ/germ-home.html) provides the structure for online discussions about reservoirs, fluxes, databases, and other scientific or technical aspects of the information in the model (Staudigel et. al 1998).

Pittman and Brown (2001) note that special collaborative tools are needed in geoscience. In paleontology, one needs the ability to post artistic and photo images of specimens, and have threaded discussions on discrete features of the image that are clickable. One needs to be able to have discussions tied to specific branches, nodes, and taxa.

Other Implications of IT for Geoscience

Workforce issues

There has been some discussion in the literature of the importance of training geologists in the use of IT. Merefield and Roche (1993), for example, note that regional groups in the British Geological Society provide training for geoscientists to use computers.

Productivity

Little literature directly addresses the affects of IT on productivity in science. Dengo (1998) notes that IT has increased productivity in petroleum exploration. Exploration companies need to manage increasing volumes of digital data. IT allows geoscientists to integrate their knowledge in ways that save time and cost while continuing to improve the quality of data. Exploration tasks that traditionally were done sequentially (seismic data acquisition through processing and interpretation) are beginning to be done in parallel. The time to do these tasks will further decrease as companies improve their use of technologies for real-time data analysis and are able to globally integrate disciplines using virtual teams.

Concept of Earth System Science

A final implication of IT for geoscience is the contribution of IT to the concept of earth system science. Earth system science views the Earth as a physical system of complex interactions among the geosphere, atmosphere, hydrosphere, and biosphere. It involves understanding the interactions of chemical, physical, and biological processes over spatial scales ranging from micrometers to the size of planetary orbits, and over time scales from milliseconds to billions of years. (Johnson, Ruzek and Kalb 1999, Bretherton 1985). This new conceptual framework for earth sciences very much depends upon remote sensing data from space, data management systems to manage and make available this data, and computer models that can synthesize the complex relationships among the various elements of the system over various temporal and spatial scales.

IT in Biosciences

Evolution of IT in Biosciences

Although information technology has been used in the biosciences for a long time, the biosciences were initially relatively modest users of IT. A 1966-1967 survey that covered use of computing in the life sciences, found that a lower percentage of life scientists than physical used computers (National Academy of Sciences, 1970). In recent years, however, the use of IT in the biosciences has expanded greatly. Lander et. al. (1991) observed that biology had been going through major changes driven by computing for the previous ten years. Since the discovery in the 1950s of how genetic information is coded in DNA, the biosciences have increasingly become much more dependent on IT. Much bioscience research now involves: deciphering the information stored in genetic sequences; understanding how these sequences codes for genes and proteins; and understanding the function and structure of these proteins. In a sense, biology has become driven by the information contained in the genetic code, and information technology is critical in generating and managing this information.

The growth in importance of information technology is exemplified by the growth of several new interdisciplinary subfields of biology. One is “bioinformatics,” defined as the application of computers, databases, and computational methods to the management and analysis of biologic information. The human genome project and genome sequencing projects in other organisms, together with new technologies that analyze gene expression patterns, have created vast amounts of data. Bioinformatics has become essential for almost every aspect of data management in modern biology (Kaminski 2000). Another subfield is “computational biology,” which uses mathematical and computational approaches, such as mathematical modeling and computational simulation techniques, to address theoretical and experimental questions in biology.

Other new subfields are the ones with the suffix “omics” – which include genomics (the study of the full DNA sequence of organisms, including sequencing, mapping and determining the function of genes); transcriptomics (RNA and gene expression); proteomics (protein expression); metabolomics (metabolites and metabolic networks); pharmacogenomics (how genetics affects individuals responses to drugs); and physiomics (physiological dynamics and functions of whole organisms). Each of these involves the combination of quantitative, experimental, and computational approaches to improve understanding of complex biological systems and processes.

Today IT plays a number of roles in bioscience:

- IT-aided instruments for data collection, such as gene sequencers and microscopes;
- databases, such as gene and protein data banks;
- data analysis, such as statistical analysis, algorithms for gene sequencing, and data mining;
- imaging and visualization, such as Magnetic Resonance Imaging (MRI) and tomography;
- modeling and simulation, such as modeling of protein folding, cells, tissues, organisms, or populations; and
- communication, such as scholarly communication, and computer aided collaboration.

Much of the bioscience can be arranged in a hierarchy, going from the molecular level to the ecosystem level, and the uses of IT vary somewhat in each level of the hierarchy. Table 4 shows the main IT applications for each level of the hierarchy.

Table 4

Level	IT Applications
Genome sequences	Automated sequencing; genome databases; software for aligning, assembling, and comparing gene sequences
Genes, proteins, RNA,	Microarrays, gene expression databases, modeling protein folding and structures,
Biochemical pathways and processes	Modeling metabolic pathways, databases of pathways
Cellular and Developmental Processes	Models of cell growth and development
Tissue and Organismal physiology	Imaging, models of organs
Ecological Processes and Populations	Population models, taxonomic databases

The following sections are organized by subfield of biosciences, starting at the molecular level and proceeding up to cells and tissues, organisms, and populations.[13]

IT and Molecular Biology

While the original gene sequencing techniques as developed by Sanger (Sanger et. al. 1977) did not involve information technology, modern gene sequencing and the subsequent efforts to interpret and apply the results of gene sequencing have depended heavily on information technology.

Information technology is used in genomics in several areas. Modern high-throughput gene sequencers are highly automated in their operation -- through the use of robotics -- and in their data collection. They produce vast amounts of raw sequence data, which is stored in digital form. Assembling the sequences of gene fragments into genomes also relies on computation to properly assemble the sequences. One of the main sequencing strategies, the "whole genome shot gun method," relies especially heavily on computation (Venter et. al 1998). In this technique, random samples of the genome are sequenced and then reassembled using computer algorithms. When applied to the sequencing of large genomes such as the human genome, the strategy involves producing and then reassembling millions of sequence fragments. In addition to the draft human genome sequence published in February 2001, genome sequencing of many other species has been completed, and hundreds more are in progress.

---

[13] This worked better than organizing the material by IT application because, in practice, many of the IT applications are not easily separated. In genomics and proteomics, for example, use extensive shared databases and data analysis tools that are available through the Internet in an integrated way. Moreover, similar sounding IT applications are quite different in different subfields. For example, the modeling of protein folding uses very different methods than modeling cells or ecosystems.

The sequence data that result from sequencing projects are stored in large, shared, Internet-connected databases. Data is both provided and retrieved via the Internet. DNA sequence data that are produced by laboratories around the world are managed by three groups: the European Molecular Biology Laboratory (EMBL), GenBank, and the DNA Databank of Japan (DDBJ). The databanks exchange data on a daily basis (Kusnirikova and Celarova 2000, Yao 2002). The GenBank contains sequences of more than 20 billion of ATGC[14] base pairs and is expanding very rapidly (www.ncbi.nih.gov/Genbank). (For current GenBank statistics, see: http://www.ncbi.nih.gov/Genbank/genbankstats.html.)

Sequencing the genome is just one step in the process of getting biologically useful information out of the genome. The greater challenge is to use gene sequences to increase understanding of biological systems. Other steps are:

1. identifying genes;
2. annotating gene sequences;
3. comparing the genomes of species;
4. determining the structure and function the proteins encoded;
5. understanding the regulation of gene expression (e.g., which genes are active in various cells and tissues at various times); and
6. determining genetic differences among individuals.

All of these steps make extensive use of information technology. They use data stored in various gene and protein databases, and use a variety of algorithms and computer tools to match gene and amino acid sequences. Determining the structure of proteins uses a variety of modeling and imaging technologies, as well as databases and search tools.

Gene Identification and Annotation

Much genomic work has centered on finding genes[15] in the genomic sequences. This can be done by comparing a sequence with sequences that have been identified as genes in other species, or by predicting genes from first principles. There are a wide variety of computational tools that are used for making these analyses (Tsoka and Ouzounis 2000, Kramer 2001, Wishart and Fortin 2001). One of the better known search tools BLAST (basic local alignment search tool), which is used for searching databases for nucleotide patterns. It allows the comparison of an unknown DNA or amino acid

---

[14] Adenine (A), thymine (T), guanine (G), and cytosine (C) are the chemical bases that make up the DNA molecules.

[15] Genes are the specific sequences of nucleotides in a particular position on a chromosome that encode a specific product (usually a protein or RNA molecule).

sequence with sequences from human or other organisms (Bayat 2002). A variety of other software packages is also used for sequence assembly and analysis, and can be used to can manipulate and annotate individual sequences and sequence elements. Genomic databases and a large number of tools for analyzing the databases are available on the Internet. The National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) maintains many of these tools (including BLAST) and databases, and provides links to others.

There are also a variety of computational tools for predicting genes based on their nucleotide sequences (Claverie 1997, 1998). The best programs (in 1997) located more than 80 percent of the internal coding exons (genetic sequences that are used in mRNA)[16]. With this level of accuracy, computational methods are useful but do not eliminate the need for experimental validation.

In addition to sequencing the human genome, another important aspect is annotating the genome – associating with each segment of the genome additional information about the segment's function or origin (Rust et. al 2002). An important step is the association of the sequence with what is known about its functional role, such as its role in metabolism. Software for genome annotation, such as Ensembl (www.ensembl.org), has become more important in gene discovery (Yao 2002). Bioinformatic tools are also needed to help effectively distribute annotation data to the bioscience community.

Comparative Genomics

Comparative genomics – comparing the genomic sequences of different organisms -- emerged as a research field after the first genomes were sequenced in 1995 (Yao 2002). The genomes of a rapidly growing number of organisms, including the bacterial pathogen *Haemophilus influenzae,* archaea, yeast, the worm *c. elegans*, the fruit fly *Drosophila melanogaster*, and mouse have been sequenced. Research on many other organisms is in progress. Comparative genomics helps to identify genes and also sheds light on evolution. Comparing genomes is computationally intensive – for example, comparing the genomes of two mammals requires comparing billions of nucleotide sequences in a two-dimensional manner (Yao 2002).

One result of comparative genomics is the Clusters of Orthologous Groups (COG) database (www.ncbi.nlm.nih.gov/COG), which identifies the gene sequences that encode for proteins that occur in the genomes of different phylogenetic lineages (Tatusov et. al. 1997). Orthologs are genes in different species that evolved from a common ancestral

---

[16] DNA sequences that get translated in messenger RNA are called exons ("expressed sequences"). The DNA sequences that are not used in mRNA are call introns ("intervening sequences.") Thus to go from a DNA sequence to a mRNA sequence that codes for a protein, the intron sections of the sequence must be removed and the exon segments must be stitched together. The fact that exons are interspersed among introns makes the tasks of identifying genes much more difficult.

gene by speciation. Normally orthologs retain the same function. Orthologs are genes that have been retained during the course of evolution, usually because they perform important functions, and the analysis of orthologs is an important tool in understanding both the function of genes and the course of evolution. As of the end of 2002, 3307 clusters of orthogonal groups had been identified, based on analyses of 43 complete genomes, representing 30 major phylogenetic lineages.[17]

Functional motifs (sequence patterns that are linked to a biological function) can be extracted by comparing sequence information on the same or similar functions of various species. The extracted functional motif database is available (http://www.expasy.ch/prosite) and can aid in the prediction of functions.

Computer aided tools have also been developed to test evolutionary hypotheses – such as which evolutionary tree fits best with observed differences in gene sequences among species (Huelsenbeck et. al. (1997).

Protein Structure

Understanding protein structure and function is a central problem in getting biologically valuable information out of genomics (Attwood and Miller (2001). As with gene sequencing, the fundamental understanding of protein folding predates the recent expansion of information technology in biology (Honig, 1999). The recent progress in protein structure prediction, however, has been due primarily to the explosive growth of sequence and structural databases, as well as advances in modeling.

Functional genomic and proteomic technologies are producing biological data relating to hundreds, or even thousands of proteins per experiment (Weir et. al. 2001). The various genome sequencing projects provide complete amino acid sequences of proteins that serve many different biological functions, including catalysts, inhibitors, messengers, transporters and structural elements of organisms (Pieper et. al. 2002).

Major non-redundant protein databases are SWISS-PROT and Protein Information Resource (PIR) (Kusnirikova and Cellarova 2000). The SWISS-PROT database consists of properly checked and annotated translations of sequences in the EMBL database. The PIR database was created by the National Center for Biotechnology Information (NCBI) as translations of the gene sequences in GenBank.

To understand the role of proteins, it is useful to know the three dimensional structure of the proteins. Only a small portion of known protein sequences have had their structure determined through X-ray crystallography or nuclear magnetic resonance spectroscopy. There are about 16,000 proteins in the Protein Data Bank that have known structure, whereas there are over 600,000 entries in the major protein sequence databases

---

[17] Data from http://www.ncbi.nlm.nih.gov/COG, accessed January 1, 2003.

(Pieper et. al. 2002). For the rest of these proteins, it is useful to predict their structure through means that rely on IT.

Two classes of computation-based methods to predict protein structure are commonly used. One type of method predicts the protein structure through modeling based on the protein's amino acid sequence alone. While this prediction method can work well for small proteins, the error rates are too great for larger proteins (Pieper et. al. 2002). The CASP worldwide competition of protein structure prediction, which has been held in every other year since 1994, has been making a great contribution to accelerate the improvement of the above-mentioned methods.[18]

A second method is to rely on detectable similarities between the sequence of the protein under investigation and sequences in other proteins with known structures. In this method, one finds a known structure (based on a protein that has had its structure determined, such as through X-ray crystallography) that is related to the sequence whose structure is to be modeled. One aligns the sequence to be modeled with the sequence whose structure is known, and then builds a model of the new structure based in part on the known structure and in part on the calculations in the model. The accuracy of the models formed in this way correlates with sequence similarities of the two proteins (Pieper et. al. 2002). Automated modeling tools are available on the Internet. MODBASE is a database of annotated comparative protein structure models.

There is a major structural genomics initiative to determine the 3D structure of all proteins (see http://www.structuralgenomics.org/).[19] This project is based on the idea that if the representative structures of basic protein folds and basic family members can be determined, then eventually almost all of the protein structures derived from genomic genes can be determined (Stevens et al., 2001). The project involves:

- organizing known protein sequences into families;
- selecting family representatives as targets;
- solving the 3D structure of the targets by X-ray crystallography or NMR spectroscopy; and
- building models for other proteins by homology to solved 3D structures.

IT has also made contributions to X-ray crystallography and NMR determination of protein structure. Augen (2002) notes that fifteen years ago, X-ray crystallography of proteins was a long and complex tedious process that took months. Diffraction data were collected on film, measurements were made by hand, analysis software was crude, and

---

[18] (Critical Assessment of Techniques for Protein Structure Prediction, http://predictioncenter.llnl.gov)

[19] See also the unauthored, undated white paper "Advanced Computational Structural Genomics." At http://cbcg.lbl.gov/ssi-csb/Meso.html

models were built by hand. Today X-ray diffraction data are fed directly into computer systems that quickly calculate candidate structures. These structures are displayed on advanced desktop systems, and researchers can easily examine the data easily by rotating and modifying the structure. Other computationally intensive structure-prediction techniques, such as nuclear magnetic resonance (NMR), have also advanced rapidly, and researchers can now combine multiple sources of information to help predict structures more accurately than was previously possible.

Attwood and Miller (2001) caution that there are still many challenges in determining the structure and function of proteins. It is not safe to propagate functional annotation from one sequence to another merely on the basis of some degree of shared similarity. They note that methods to predict structure are still unreliable, and, due to the degree of automation that has necessarily taken place with imperfect tools and protocols, there is an unknown quantity of misinformation in the databases upon which further analyses are based.

Gene Expression

Another major area of work is analyzing gene expression – determining which genes are active in which tissues and at what time. There are a variety of technologies, such as the DNA microarrays[20], protein chips[21], 2D-PAGE[22], Yeast two-hybrid[23] that are being used to develop gene expression and protein interaction data. They are used to determine which proteins are active at what times, and this helps determine the function of proteins.

DNA microarrays, for example can rapidly provide a detailed view of the simultaneous expression of entire genomes and can provide insight into gene function and disease pathology (Greenberg 2001). As with gene sequencing technologies, these technologies are producing huge quantities of data that are processed using a variety of computational tools and are stored in large databases that are made available on the Internet. Databases such as the gene expression database (http://genome-www4.stanford.edu) and protein interaction database (http://dip.doe-mbi.ucla.edu), as well as many others are now available via the Internet. Developing new ways to extract useful information from these data sets is a new challenge for bioinformatics.

---

[20] In DNA microarrays, a large number (thousands or 10,000s) of known oligonucleotides or cDNA are arrayed on a chip or glass slide. Flourescently labled target DNA or RNA from a sample mixture bind to the matching sites on the array. The presence or absence in the sample of thousands of specific nucleotide sequences can be determine by the patterns of fluorescent emission.

[21] Protein chips are similar to DNA microarrays, but array proteins, rather than nucleotides, on the chip. They are a less mature technology than DNA microarrays.

[22] 2D PAGE stands for "two-dimensional polyacrylamide gel electrophoresis."

[23] The yeast two-hybrid system is a method for detecting protein-protein interactions in vivo.

Another approach to study gene expression is to clone and sequence the full-length cDNAs[24] from cells of interest (Yao 2002). This cDNA information is very useful for the analysis and annotation of genomes.

Weinstein (2001) notes that with 35,000 genes and hundreds of thousands of protein states to identify, correlate, and understand, we have entered the "omic" (genomic, proteomic, etc.) era in biology. “Omic” research is a different approach from the more traditional study of one gene, gene product, or process at a time (Weinstein 1998). Often, one generates a database of molecular information with only limited ability to predict what about it will prove most useful. Large-scale omic studies of cellular molecules in aggregate rarely answer interesting questions, however, without the assistance of information from traditional hypothesis-driven research. The two types of science are synergistic.

Applications

Genomics and bioinformatics are leading to many applications in biomedical research (Yao 2002). One example is “genome-based drug discovery,” which is expected to increase the number of drug targets dramatically (Yao 2002). Several thousand genes related to diseases may be identified in the near future. Also, hundreds of pathogenic microbial genomes, each of which has hundreds or thousands of genes will be identified. These achievements will help increase the number of drug targets. Duckworth and Sanseau (2002) note that there are now many examples in which pharmaceutical companies have identified genes of interest initially by *in silico*[25] analysis.

Another concept is “structure-based drug design” (Sun and Cohen 1993, Yao 2002, Klebe 2000, Zeng 2000), which has emerged as an alternative to the traditional process of drug discovery through screening. In structure based drug design, the focus is on understanding protein structure and the relation between proteins and other molecules that bind to the protein (ligands).[26] If the three-dimensional structure of a given protein is known, this information can be used to choose or improve existing ligands or design new ligands. Computational methods supplemented by molecular graphics are used in this step. The features of the protein’s receptor area can be used in queries for computer-based screening of large compound libraries to identify promising ligands. These ideas must then be confirmed experimentally. Subsequently, further work goes on to optimize the ligands for higher affinity and better selectivity.

Another emerging application of genomic information in medicine is personalized medicines and medical treatments based on the genetic variations of an individual. The

---

[24] cDNA (complementary DNA) is single-stranded DNA that is complementary to messenger RNA. It shows the genes that are active in a cell.

[25] *In silico* refers to doing experiments or analyses in the computer, rather than in living organisms (*in vivo*) or in the laboratory (*in vitro*).

[26] A ligand is an atom, ion or functional group that is bonded to one or more central ion(s) forming a complex.

study of individual differences by means such as Single Nucleotide Polymorphism (SNP) is widely expected to be a promising approach in identifying useful genetic markers that could be employed in the analysis of drug responses and the detection of disease-associated genes. SNP analysis looks for single nucleotide variations in an individual's genome. SNPs are genetic flags that are often linked to susceptibility to diseases or unusual responses to drugs. SNPs have led to discoveries about the genetics of diseases such as sickle-cell anemia and Down's syndrome, and they have allowed testing for predisposition to disease such as breast cancer and cystic fibrosis. A variety of computer programs are used to identify SNPs. In addition, studies linking SNPs to disease/non-disease group data are highly dependent on statistical or mathematical methods.

There are a wide variety of databases and other Internet resources related to mutations (Scriver and Nowacki 1999, Claustres et. al. 2002). Some are organized by disease while others are organized by chromosome or by ethnic group. There is now a professional society (The Human Genome Variation Society), and a journal (*Human Mutation*) devoted to this. A large number of mutation databases are listed at http://www.genomic.unimelb.edu.au/mdi/dblist/dblist.html

There are a wide variety of other genomic databases related to diseases. Some examples of these are:

- Schaefer et. al. (2001) describes the National Cancer Institute's Cancer Genome Anatomy Project (CGAP), which is developing publicly accessible information, technology, and material resources for cancer-related genomics. The project focuses on building and annotating catalogs of genes expressed during cancer development; identifying polymorphisms in those genes; and developing resources for characterizing cancer-related chromosomal aberrations.

- Huret et. al (2000) describe the "Atlas of Genetics and Cytogenetics in Oncology and Haematology" (URL: http://www.infobiogen.fr/services/chromcancer), which is a database devoted to chromosome abnormalities in cancer, cancer-prone diseases, and genes involved in cancer.

- Oh et. al. (2001) developed a gene expression database to identify molecular changes in lung cancer. A large number of proteins that are expressed in different types of lung cancer have been identified. The database is intended to help develop classification schemes for lung cancer and identify markers for early diagnosis.

- Barnes (2002) describes database and data mining used to identify genes associated with psychiatric diseases.

Genomics is also being used in non-medical application, including zoology, microbiology and plant biology. The complete catalogue of plant genes and proteins, revealed by genome sequencing, has great potential to provide insights into cell biology (Bevan 2002). The same analytical methods used to analyze genome sequence data, including analysis of the transcriptome, proteome and metabolites can be applied to plant biology. An example of this is the MaizeDB (http://www.agron.missouri.edu/), which is a public Internet gateway to current knowledge about the maize genome and its expression (Polacco et. al. 2002). Another broadly based collaboration is the transatlantic yeast gene deletion project, a collaboration involving researchers in the USA, Canada and Europe (Kelly et. al, 2001.)

IT and Biological Systems

Beyond genomics, IT is having a substantial impact on biology at scales above the molecular level. Some of this work is under the rubric of "systems biology." The living body is composed of numerous subsystems, by which the flows of energy, material and information are controlled. Elements of this include genetic regulatory systems, metabolic systems, cell cycles, various physiological and pathological systems, organ systems, and other systems from the molecular level, cellular level, tissue level, and organ level to the organism level (Yao 2002). Unlike traditional biology, which typically examines single genes or proteins in isolation, systems biology simultaneously studies the complex interaction of gene, protein, and cell elements that form informational networks and systems.

Systems biology aims to model and simulate various systems and visualize the results to help better understand living processes. Instead of the data driven approach of genomics, research in biological systems can be characterized as “model-driven’’ (Yao 2002). The complexity of biological systems and the explosion of the quantity of biological information that is rapidly becoming available from experimental and clinical studies require the use of mathematical and computational modeling (Mehr 2001). These systems are too intricate to study without advanced computational tools for managing and integrating the data into mechanistic models (Department of Energy 2001).

Modeling requires taking knowledge in the form of qualitative biological theories and expressing it as explicitly and quantitatively as possible. It requires that implicit knowledge be made explicit so that disparate human knowledge can accumulate in an integrated way (Yao 2002). Model building also helps point out areas where knowledge is lacking, such as unknown pathways and parameters. Simulation can identify missing components and suggest experiments to gather missing information.

The ultimate goal of this approach is to develop a ‘‘Life Simulator’’ (Yao 2002). The path toward this is to develop, step by step, subsystem simulators of subcellular mechanisms, whole cell simulators, cell development simulators, organ simulators, physiological simulators, pathological simulators, and body simulators. There are many systems biology projects underway including models at the cellular level, tissue level, and organism level.

Cellular and Subcellular Levels

There are a wide variety of activities going on to apply information technology to model activity at the cellular and subcellular levels. Work is intended to produce both fundamental advances in science as well as practical applications in medicine and agriculture. This work is expected to eventually enable the drawing of the entire metabolic map of a cell.

Determining the biological function of genes, and understanding how they interact to yield a living cell, is a major challenge of the post genome-sequencing era (Van Helden et. al. 2000). Understanding is needed of the functional pathways in a given cell or tissue, including processes such as metabolism, gene regulation,[27] transport, and signal transduction.[28]

Understanding the regulation of gene expression is key to understanding the functioning of organisms at a molecular level (de Jong 2002). Technologies such as DNA microarrays along with computer-aided algorithms are being used to analyze the behavior of thousands of genes at a time (see previous discussion of gene expression), and are creating a foundation of data for building integrated models of cellular processes (Schilling et. al 1999 and Kao 1999). Gene expression is regulated through networks of interactions between DNA, RNA, proteins, and small molecules. Because most genetic regulatory networks involve many components connected through interlocking positive and negative feedback loops, it is difficult to obtain an intuitive understanding of their dynamics. Formal methods and computer tools for the modeling and simulation of genetic regulatory networks are needed to understand their operation. Although past attempts to model cellular processes have been disappointing, increases in biological understanding combined with advances in computational methods and in computer power make it possible to foresee construction of useful and predictive simulations of cellular processes (Endy and Brent 2001, Covert et. al. 2001). Computer simulations can also provide an accurate qualitative description of the signal transduction processes in plants, and can represent the dynamics of the signalling network (Genuod et. al 2001).

There are many major projects underway that are developing database or models related to cellular metabolism:

- The Alliance for Cellular Signaling (AFCS) project (www.afcs.org) is examining the mechanism of the signal transduction system inside cells (Yao 2002).

- The Consortium for Functional Glycomics (http://glycomics.scripps.edu/) is defining the how carbohydrate binding proteins function in cellular communication. The project is developing and maintaining glycomics databases

---

[27] Genes regulation refers to controlling how when and at what level genes are expressed.

[28] Signal transduction refers to how cells communicate with their extracellular environment. This happens through the interaction of cellular receptors with signals originating from other cells or from the extracellular matrix.

with Internet-based interfaces for disseminating data to the participating investigators and the public.

- The "E-cell simulation system" (www.e-cell.org) is aimed at whole cell simulation of model organisms, such as Escherichia coli and Bacillus subtilis (Yao 2002). E-cell is a software environment for building models of the cell that incorporate gene regulation, metabolism and signaling, including functions of proteins, protein-protein interactions, protein-DNA interactions, regulation of gene expression and other features of cellular metabolism (Butler 1999).

- The Cell Migration Consortium (www.cellmigration.org) is focused on research on cell migration (movement of cells within organisms). Computer models are helping researchers design experiments, test hypotheses, and integrate data (Chicurel 2002).

- The Japanese Ministry of Agriculture and Forestry is funding a ‘‘Rice Genome Simulator" to simulate the workings of the rice plant (Yao 2002).

- There are a wide variety of photosynthesis-related web sites (Orr and Govindjee 2001).

An active area of systems biology is modeling the complex interactions between various metabolic systems and the effects of different pharmaceutical compounds on those systems (Augen 2002). This approach has begun to yield insights into the effects that occur when a complex metabolic system is perturbed at the molecular level. This is expected to aid drug discovery. *In silico* models are also used for predicting absorption, distribution, metabolism and elimination (ADME) properties for candidate molecules in drug development (Butina et. al. 2002, Ekins and Wrighton 2001). A variety of types of simulations are also useful tools in understanding metabolism, predicting drug-drug interactions, and other pharmacokinetic parameters.

Tissues/Organs

There are a variety of projects simulating biological systems at the tissue/organ level. One prominent project is the Virtual Heart Project at Oxford University, UK, which has constructed a huge model of the heart mechanism with the collaboration of more than 80 international researchers (Yao 2002, New Scientist 1999). The model contains more than 1 million cells or elements, each of which has internal complex biochemical reactions, and the model is governed by more than 30 million equations in total.

Similar approaches are underway on many organs, such as lungs, pancreas and kidney, as well as the immune systems (Yao 2002, Mehr 2001). Some of this work is coordinated as part of the international Physiome Project (www.physiome.org). Other researchers have been developing models of vision (Naisberg 2001), spinal cord stimulation (Holsheimer 1998), and the hippocampal region of the brain (Allen et. al. 2001).

Computer models and simulations have been applied to a wide range of tissue/organ level problems, including:

- simulations of the uptake, accumulation and retention of drugs in tumors (Jang et. al. 2001);
- simulations of neurones firing patterns (Perez and Carlen 2000);
- simulations of neural activity (Schmajuk et. al. 2001);
- models to describe antibody-antigen binding and the structural models of receptors in immunology (Merrill 1998);
- models of water transport systems in land plants (Roth et. al. 1998);
- models of the spatial structure and development of plants (Prusinkiewicz 1998);
- cerebral circulation computer models (Charbel et. al. (1998); and
- models to simulate traumatic brain injury (Bandak 1995).

Organisms

Modeling all levels of biological complexity is beyond the capabilities of current computers, but each increment in computing makes it possible to move up the biological complexity ladder. Most models at the level of organisms and tissues do not go all the way back to the genetic/molecular basis. There is, however, for the first time the potential to use large-scale computation to simulate the behavior of living organisms as whole complex dynamic systems.

There are a variety of projects aimed at modeling whole organisms. One is the virtual patient system "PhysioLab (www.entelos.com) which includes a virtual obesity model, diabetes model and asthma model (Yao 2002). There are a also large number of databases containing information on a wide variety of model organisms used in scientific study. Some of the species for which large databases have been established include many yeast, bacteria and archea, arabidopsis (plant), fruit flies, zebra fish, blowfish, *c. elegans* worm, and mice.. Lists of many of these databases are available at Oak Ridge National Laboratory's Virtual Library of Genetics (http://www.ornl.gov/hgmis/vl_organisms.html).

Populations

Computer statistical methods, databases, and modeling, and have been used for a long time in population biology and ecology. A variety of statistical methods have been used to help address problems associated with nonrandom sampling, unknown population distributions, heterogeneous variances, small sample sizes, and missing data (Pitt and Kreutzweiser 1998). Taxonomic databases have been developed on many different species. For example, Poulin et. al. (2001) discuss computerized databases on diatoms, and suggest that such databases aid transmission of data to colleagues around the world and constitute a major advance in the field.

Bioinformatics and the Internet are playing a large role in biodiversity science (Bisby 2000). Biodiversity science by its nature involves scientists from around the world and databases on taxonomy and animal and plant records. The Internet and bioinformatics are especially useful in bringing these together. A major issue is achieving interoperability among the distributed databases (Edwards et. al. 2000).

Geographical information systems (GIS) are also important in population and environmental biology, due to the spatial nature of much of the data. GIS, especially when combined with statistics, provide a powerful set of tools for spatial analysis in the agricultural, earth and environmental sciences (Burrough 2001). GIS systems originally concentrated on automated map making and facilitating the comparison of data on maps. Recently, standard statistical packages have been linked to GIS for exploratory data analysis, statistical analysis, and hypothesis testing.

Population models are used for a variety of purposes and scales. IT has allowed the integration of data, model creation and testing of diverse and complex population interactions among biospheric and hydroclimatic systems, both spatially and temporally. For example, Rogers and Johnson (1998) investigate approaches to modeling the regeneration of oak dominated forests. Loehle and Leblanc (1996) describe forest simulation models to predict changes in forest composition, forest dieback, or loss of forest cover in response to increased temperatures associated with increasing atmospheric carbon dioxide concentrations. Cabeza and Moilanen (2001) analyze models used to help identify sets of nature reserves that maximize the representation of regional diversity.

Rogers and Johnson (1998) note that process in forests are highly complex, with a very large number of interactions among highly variable biotic and abiotic factors. They view models as a tentative description of a system or theory that accounts for all of its known or important properties. Models range along a continuum of complexity from very simple to extremely detailed, and they can be theoretically or empirically based. The more complex computer models may 'simulate' real phenomena.

One approach to modeling ecosystems complexity is “individual-based models,” which fall under the new field of complex adaptive systems (Railsback 2001). Such models study how complex behaviors emerge in systems of relatively simple interacting individuals. Some key issues in such models are:

- choosing what behaviors and population dynamics should emerge from the model's key processes vs. being imposed on the model;

- determining how individual traits should be modeled so that realistic population responses emerge; and

- choosing the adaptive processes of individuals to be modeled and the measures of fitness that are appropriate to use as the basis for modeling decision making.

Sigmund (1998) notes that a wide variety of computer simulations show that cooperation can evolve in populations of selfish agents, both with direct and indirect reciprocation. An example is computer simulations of host-parasite coevolution (Little 2002). *In silico* experiments have shown that host-parasite interactions could be responsible for high levels of genetic diversity in host populations, and even be the principle determinant of rates of genetic recombination.

Several people urge caution in using computer models in making practical decisions. Johnson (1995) notes that fishery management models should be used for quantitative predictions only after all model inputs have been estimated from empirical data and the model has been tested for agreement with an independent data set. Models can be more useful as tools for organizing data and concepts, learning about the system to be managed, and exploring management options. Models can provide valuable input to the decision-making process and can help determine priorities for data collection.

Moir and Mowrer (1995) note that forest management decisions are often based on long-term projections from computer models. Although these models are theoretically based and statistically calibrated, they usually fail to account for uncertainty in the underlying assumptions, in the statistical calibration, and in the values used to initiate projections. Moreover, given current knowledge of ecosystem behavior, even the best models may not capture important features of the ecosystem.

Cross Cutting Topics

A number of topics cut across the various levels of biosciences that served as the organizing basis for the previous sections. These include literature about IT applications that affect several levels of bioscience, as well as some conceptual approaches that do not fit neatly in one area. These include instrumentation and imaging, electronic communication and collaboration, and the fields of artificial intelligence and artificial life.

Imaging and Instruments

IT-aided advances in imaging and instrumentation have had a major influence in the biosciences. Techniques have been developed for digital processing of microscope imaging, allowing computer-based image enhancement by filtering techniques, deblurring, and contrast enhancement (Sabri et. al. 1997). Computer-based methodologies have been developed to combine the advantages of scanning electron microscopy and conventional histology with tissue recognition (Kaufman et. al. 1998). With these technologies, tissues can be highlighted in a particular color and viewed either in isolation or in combination with other appropriately labelled tissues and organs. Tissues can be shown in any orientation as a transparent overlay on computer-generated histological sections or as 3-D images. In addition, several systems for telemicroscopy (remote operation and/or viewing of microscopes over the Internet) have been developed, allowing remote consultation or collaboration (Petersen et. al. 2000).

Work has been ongoing to develop three-dimensional visualization of many tissues and organisms. Whiten et. al. (1998) have devised techniques for the production of interactive 3D models reconstructed from serial histological sections of human embryos, focusing on developmental changes occurring in embryogenesis. Sarwal and Dhawan (2001) discuss three-dimensional (3-D) reconstruction of coronary arterial trees that allows clinicians to visualize vascular geometry. Three-dimensional images are being combined into large information systems in anatomy that can be used by clinicians, researchers, educators, or students. Examples include create 3-D atlases of the brain, knee and thorax (Brinkley et. al.1998). These on-line atlases get extensive use by researchers and students around the world. Another large project is a digital atlas of mouse development, which covers mouse developmental anatomy from fertilization to birth (Kaufman et. al. 1998).

Artificial Intelligence

Another cross-cutting area at the intersection of IT and bioscience is the application of artificial intelligence to biology. Molecular biologists and computer scientists have experimented with various computational methods developed in artificial intelligence, including knowledge-based and expert systems, qualitative simulation, and artificial neural networks and other automated learning techniques (Rawlings and Fox 1994). These methods have been applied to problems in data analysis, construction of databases with advanced retrieval capabilities, and modeling of biological systems. Practical results have been obtained in identifying active genes in genomic sequences, assembling physical and genetic maps, and predicting protein structure. Artificial neural networks have many applications in biology and medicine (Dayhoff and DeLeo 2001), including extracting previously unobserved or unrecognized phenomena within the raw data, and building computer models that are useful for medical decision-making. They can be used to help predict outcomes for individual patients (rather than simply predicting patient outcome based on their statistical group).

Collaboration and Communication

Information technology is also used to enable new forms of collaboration among geographically separate researchers. Arzberger and Finholt (2002) discuss the results of past “collaboratories” -- network-based virtual laboratories -- and issues for future electronic collaboration in the bioscience. There have been a number of collaboratory projects in the biosciences, including the BioCoRE Collaboratory for Structural Biology (University of Illinois), the Microstructure Image-based Collaboratory (San Diego Supercomputing Center), and the Great Lakes Regional Center for AIDS Research. Over the past decade collaboratories have evolved from an initial focus on remote operation of instruments to a broader focus on data integration, data access, and tools to support collaboration with data.

Arzberger and Finholt (2002) note that biomedical research increasingly requires the integration of "cross scale data" -- data at the molecular, cellular, and organism levels. This requires collaboration across labs that specialize in these different kinds of data and expertise from specialists with diverse training. Collaboration is driven both by the need to share data and to share knowledge about data. These needs, along with the dramatic growth in the capacity to produce and transport data, are leading to science being reorganized around the free availability and flow of data at unprecedented volumes and detail. They suggest that key capabilities for successful future collaboratories will be:

- communications and resource control – including teleoperation and teleobservation;

- information sharing – which includes the creation and curation of data repositories, security and authentication controls, and tools for collaborative visualization and analysis of data;

- coordination – which includes planning experiments and computer runs, and scheduling scarce resources; and

- technology development – such as ensuring that new collaboration technologies are compatible with emerging standards.

Other issues related to electronic communication and collaboration discussed in the bioscience literature include the challenges of managing the increasing volume, complexity and specialization of knowledge expressed in this literature (Mack and Hehenberger 2002) and role of electronic scholarly communication (Schoonbaert 2000).

Artificial Life

Another aspect of the intersection of IT and the biosciences is the field of artificial life. This includes work in a branch of computer animation to create active self-powered objects living artificial lives in the theoretical biology zone (Hokkanen 1999). These works include animated simulations of legged locomotion, flexible-bodied animals swimming and crawling, artificial fish in virtual ecosystems, automated learning of swimming and the evolution of virtual creatures with respect to morphology, locomotion and behaviour. Artificial life also includes work on self-replication over the past 50 years (Moshe 1998). Due to their self-replicating nature, computer viruses may be viewed as a form of artificial life (Spafford 1994).

Robot-based research may also be a new methodology for biology (Webb 2000). Robots can be used as models of specific animal systems to test hypotheses regarding the control of behavior. This approach has led to novel hypotheses for animal behavior.

Workforce Issues

Butler (1999) noted that principal limit in the development of advanced computing in biology is the lack of biologists who know how to do it. The National Institutes of Health Working Group on Biomedical Computing (1999) observed that many researchers who 5 years ago (before 1999) spent little time at the computer report that they now spend 90 percent of their research time in front of their monitors. They noted that the transformation in biology requires new skills. They also noted that one needs computer expertise without sacrificing biological expertise, and this requires more team-based work.

Conclusion

While the lack of controlled evaluations in the literature reviewed precludes strong conclusions about the role of IT in the geosciences and biosciences, a series of observations emerge from the review.

IT is used in a wide variety of applications in both geoscience and bioscience. These applications include automated data collection, statistical analysis of data, Internet-accessible shared databases, modeling and simulation, imaging and visualization of data and analysis, Internet-based communication among scientists, and electronic dissemination of research results.

IT plays a role throughout the scientific cycle in both fields. IT is used in many ways throughout the scientific process. This distinguishes IT from most other kinds of scientific equipment, such as microscopes or physics accelerators, which are generally used in one stage of the scientific process, such as data collection. IT aids in hypothesis formation, research design, data collection, data analysis, and communications of scientific results. Table 5 shows the relation of IT to stages of the scientific process.

Table 5. Relation of IT applications to stages of scientific process

Stage of Scientific Process	**IT Application**
Hypothesis formation	Data mining, modeling
Research design	Modeling and simulation
Data collection	IT-aided instrumentation, databases
Data analysis	Statistical packages, modeling, simulation, imaging
Information dissemination	Websites, electronic journals, databases, maps
All stages	Email, other communication and collaborative technologies

Data mining can help identify patterns that aid in formulating new hypotheses. Modeling often helps identify where better data is needed. IT-enabled instruments aid in collection of data. Internet-connected databases are used to store data and to allow the global scientific community to access and often to contribute to the database. Modeling, simulation and visualization are used to analyze and synthesize the data, and to aid in the understanding of complex relationships. Electronic publications, maps, and digital libraries aid in disseminating the results. Some technologies play roles in a variety of parts of this process. Electronic communication technologies aid communication among scientists throughout the process. Shared databases also serve to aid communication among scientists and disseminate results. Modeling is used to explore hypotheses, identify data needs, and interpret data.

IT is associated with major changes in both fields. IT is associated with the development of new subdisciplines in each field, as represented by the development of professional societies and specialized journals, such as *Computers & Geoscience, Bioinformatics* and several other journals. More importantly, IT appears to be contributing to major conceptual changes in each of the two fields. IT is playing an especially important part in a transformation in the biosciences, based on the understanding that biology is at its core an informational science -- based on the information embedded in the genetic code. A key challenge in biology is to understand the instructions encoded in the genomes, such as the structure and function of the proteins, and the regulation and expression of genes. IT is essential to store, manage, and decipher the mass of information produced in this work. At levels of organization above the molecular level, computer models are critical tools used to handle the complexity of the relationships in biology. Computer models are being created of cells, tissues, organisms, and populations.

In geoscience, the changes have been more gradual -- IT has been used for decades in the geosciences, especially in geophysics and in petroleum exploration. More recently, IT has played a part in the emergence of "earth system science" as an integrative view of the geological, oceanographic, atmospheric and environmental sciences. This concept was largely driven by the availability of space-based remote sensing, but also relies on large shared databases and modeling of large-scale earth processes. Modeling has led a shift from geology being an observational and descriptive science to one that is more predictive although, the use of model-based predictions in public policy, while essential for critical decisions, is often controversial. IT has also led to a change in the representation of geologic information, from paper maps to digital databases and geographic information systems.

IT-enabled science and traditional science are frequently complementary and synergistic in the two fields. The fundamental techniques in gene sequencing and gene expression detection were not based on IT, but IT enabled enormous leaps in the efficiency of data collection. Both large-scale data collection, such as gene sequencing, and hypothesis-driven small-scale studies are needed to make conceptual advances. Although modeling and simulation in some cases reduce the need for experimentation, often they identify the need for new data that must be obtained through experimentation.

There are strong similarities but also substantial differences in the way IT is used in the two fields. Each field uses similar basic information technologies, including IT-aided instrumentation, databases, modeling and simulation, electronic communication, data mining. In both fields, IT has aided more efficient data collection, is a critical tool for handling complexity, has aided greater dissemination of information, and appears to have facilitated widespread collaboration around large projects, often centered on contributions to shared databases or models of complex systems.

On the other hand the role of IT in biology is unique because information contained in the genetic code underlies much of biology. IT in geology is unusual because much of geoscience data is geospatial in nature and requires the use of geographic information systems. Modeling plays a different role in the two fields because in bioscience it is often possible to test models with experiments whereas the temporal and spatial scales involved in geology often make such testing impossible or impractical.

IT plays a particularly important role in specific areas. In biology, IT is especially important in bioinformatics and the "omics" -- (genomics, proteomics, transcriptomics, etc.). In the geosciences, IT is especially important in analyzing subsurface formations, mapping, and modeling complex systems. There appear to be few, if any, fields that are untouched by IT. Shared Internet accessible databases are important in paleontology; and models and databases are important in population biology and ecology; and genomics are influencing many fields in biology.

IT appears to influence scientific processes in several ways. These include:

- The Human Genome Project and other large scale genomics and “omic” studies have initiated a new approach to biology focused on gathering information and placing it in a database, in contrast to the traditional scientific process of using hypotheses to drive data collection. This approach is sometimes termed "discovery" or "hypothesis generating" science.

- Modeling and large shared databases facilitate work on large, systems-level complex problems (as opposed to reductionist science). Systems biology and earth systems science rely on modeling of complex systems and large databases. In contrast to “omic” studies, modeling requires the formalization of explicit hypotheses about biological process, and thus model-driven research is strongly hypothesis driven.

- Computers have enabled fields such as artificial life that have influenced conceptual thinking in the geosciences and biosciences.

The geosciences and biosciences rely on both general and field-specific IT. There are many examples of both general IT tools (such as the Internet, standard database and statistical packages, and electronic journals) as well as specialized field-specific IT, such as software for comparing gene sequences or modeling protein folding, or geospatial databases for geologic mapping.

There are a variety of policy issues that relate to the use of IT in the geosciences and biosciences. One issue is the role of models in policy decisions, for example, regarding climate change, the adequacy of nuclear waste sites, or the spread of infectious diseases. Although these models may provide the best basis for making critical decisions with incomplete information, they cannot be fully validated by experiments because of the temporal and spatial scale of the experiments that would be required, and there is debate over the appropriate use of these models. Other issues include protecting intellectual property (data and ideas); determining appropriate levels of database security; developing metadata and standards necessary for interoperability and to aid searching among different databases; and expanding IT training in science.

As mentioned above, there have been few controlled evaluations or critical analysis of the role of IT in the geosciences and biosciences. While there are studies of the use in science of specific aspects of IT (such as electronic scholarly communication and IT-enabled collaboration) there is little literature that analyzes the role of IT in changing:

- research methods (e.g., does IT contribute to more hypothesis generating research rather than hypothesis driven research? does IT lead to more modeling and simulation rather than experimentation?);

- research questions and directions (e.g., does IT enable more systems-level work rather than reductionist science?);

- productivity (e.g. does IT increase the productivity or cost-effectiveness of science?);

- the quality of science (e.g., is IT-enabled science more or less highly cited than non-IT enabled science?); and

- the organization of science (e.g. does IT promote large scale collaboration? More interdisciplinary work? Changes in the sociology of science? Changes in scientific institutions?).

There is little analytic work addressing these questions, at least with respect to the geoscience or biosciences. This suggests a large agenda for future work.

References

Geoscience References

Adkins-Heljeson, D. and T. R. Carr. 2001. "Using the Internet to Provide a Digital Petroleum Atlas for Kansas." *Annual Meeting Expanded Abstracts - American Association of Petroleum Geologists, vol.2001*: 2.

Ahern, T. and D. Simpson. 1995. “Exploiting the Global Internet for Distribution of Global Seismological Data.” *Society of Exploration Geophysicists SEG Annual Meeting Expanded Technical Program Abstracts with Biographies, vol.65I*: 1541-1542.

Amenta, R.V. 2001. “Three-Dimensional Computer Modeling of Fabric Evolution in Igneous Rocks.” *Computers & Geosciences,* 27, No. 4: 477-483.

American Geological Institute. 1997. *National Geoscience Data Repository System; Phase 2.* Alexandria, VA (DOE/BC/14833-T1, 11 Jul 1997), 120 pp.

Andersen, A., R.E. Chapman, J. Dickman, and K. Hand. 2001. “Using Rapid Prototyping Technology in Vertebrate Paleontology”. *Journal of Vertebrate Paleontology* 21*,* No.3, Suppl.: 28.

Anderson, W. H, T. N. Sparks, J.A. Patton, X.Y.Yang, and R.E. Sergeant. 1999. “Integration of Relational Geologic Databases and a Spatial Map Database in Kentucky.” In Soller, D. ed. *Digital Mapping Techniques '99;Workshop Proceedings.* Reston, VA: U. S. Geological Survey (Report: OF 99-0386):123-126.

Applegate, D. 1998. “Policy Issues Affect Access to Scientific Data.” *Geological Society of America, 1998 annual meeting. Abstracts with Programs,* 30.

Argast, S. and T. Corey. 1998. “Using The World Wide Web For Interactive Control Of An X-Ray Diffractometer.” *Computers & Geosciences* 24, No. 7: 633-640.

Bandy, S. and D. Wallace. 2001. “GeoBASE.WA; Geoscientific Data On-line in WA.” *Cartography* 30*,* No.2 (Dec. 2001): 31-39.

Birken, R. and R. Versteeg. 2000. “Use of Four-Dimensional Ground Penetrating Radar and Advanced Visualization Methods to Determine Subsurface Fluid Migration.” *Journal of Applied Geophysics* 43*,* No. 2-4: 215-226.

Black, R and J.D. Walker. 2001. “Development and Use of a Laptop-Based Geological Mapping System: Experiences at the University of Kansas.” In Soller, D.R. ed. *Digital mapping techniques '01; workshop proceedings.* Reston, VA: U. S.

Geological Survey (Report OF 01-0223): pp.127-131, 2001. Available at http://pubs.usgs.gov/of/2001/of01-223/.

Borgmann, A. 1998. “The Transparency and Contingency of the Earth.” *Geological Society of America, 1998 annual meeting. Abstracts with Programs* 30, No.7: 202.

Boulter, M.C. 1999. “The Plant Fossil Record on the Internet.” In Jones, T. P., and N. P. Rowe (eds), *Fossil Plants and Spores; Modern Techniques.* London: Geological Society: 177-178.

Brandle, J L, N. Nishiwaki, G. Nagy, and J. Nan. 1997. “Global Database for Petrology under Internet Environment.” *Proceedings of IAMG '97, the Third annual conference of the International Association for Mathematical Geology* 3: 1027-1032.

Bretherton, F. P. 1985. “Earth System Science and Remote Sensing.” *Proceedings of the IEEE* 73, No. 6: 1118-1127.

Brice, W.R. 2000. “Geological Models; Then and Now.” *Abstracts with Programs - Geological Society of America* 32, No.7: 87-88.

Brodaric, B. 2000. “Digital Geologic Knowledge: From the Field to the Map to the Internet.” *Digital Mapping Techniques '00: Workshop Proceedings.* Reston, VA: U. S. Geological Survey (Report OF 00-0325): 3-11.

Browne, D G and J.D. Love. 1997. “More than Books and Journals: A Plea for Greater Inclusiveness in Defining ‘Geoscience Information.’” *Geological Society of America, 1997 Annual Meeting. Abstracts with Programs* 29, No.6: 173.

Buchanan, R. C. and T.R. Carr. 1997. “The Impact of Electronic Dissemination: The Experience of a State Geological Survey.” *Geological Society of America, 1997 Annual Meeting, Abstracts with Programs* 29, No.6: 172.

Buchanan, R. and T. Carr. 1998. “The Impact of Electronic Dissemination: What Do Users Think?” *Computers & Geosciences* 24, No.6: 595-596.

Burns, S. L. 1997. “Secure Websites Facilitate Geoscientific Collaboration, Progress and Technology Transfer.” *Abstracts with Programs - Geological Society of America* 29, No.6: 265.

Cajka, M.G. and S. Halchuk. 1998. “Collecting Intensity Data via the Internet: the Cap-Rouge, Quebec Earthquake.” *Seismological Research Letters* 69, No.6: 585-587.

Carr, T. R, D. Adkins-Heljeson, R.C. Buchanan, and T.D. Mettille. 1996. “The Impact of the Internet on the Public Sector: Providing Natural Resource Information and Technology to Kansas.” *Abstracts with Programs - Geological Society of America* 28, No.7: 285.

Cartwright, D. 1998. “What's Next on the Internet? A Forward Look.” *Computers & Geosciences* 24, No.7: 627-632.

Chew, K J. 1995. “Data Modelling a General Purpose Petroleum Geological Database.” In Giles, J.R.A. ed. *Geological data management.* London: The Geological Society (Geological Society Special Publications, vol.97):13-23.

Chirico, P. G. 1999. “The Development of a Digital Surface and Subsurface Geologic Map Database of Charleston County, South Carolina.” In Soller, David, ed. *Digital Mapping Techniques '99: Workshop Proceedings.* Reston, VA: U. S. Geological Survey (Report OF 99-0386): 187-193.

Coats, J.S. and J.R. Harris. 1995. “Database Design in Geochemistry: BGS Experience. In Giles, J.R.A. ed. *Geological data management.* London: The Geological Society (Geological Society Special Publications, vol.97,): 25-32.

Collins, D. R. 1998. “Visualization of Subsurface Geology from Wireline Logs.” In Soller, David, ed. *Digital Mapping Techniques '98: Workshop Proceedings.* Reston, VA: U.S. Geological Survey (Report OF 98-0487): 61-66.

Collins, L.S. and K. W. Kaufmann. 1997. “Data Model for a Paleobiotic Survey.” *Abstracts with Programs - Geological Society of America* 29 No.6: 402-403.

Committee for the Future of the U. S. National Strong-Motion Program. 1997. *Future of the US National Strong-Motion Program.* Reston, VA: U.S. Geological Survey (Report OF 97-0530): 29 pp.

Costello, R. 1998. “The Impact of New Geoscience Technology on Growing Heavy Oil Production at Cold Lake, Alberta, Canada.” *AAPG Bulletin* 82, No.10: 1904.

Davis, A.M., J.Reddy, and C. S. Southworth. 2000. “A Washington, DC Area Geologic Map Database: Working Toward Implementation of the Digital Geologic Map Data Model.” In Soller, David, ed. *Digital Mapping Techniques '00: Workshop Proceedings.* Reston, VA: U. S. Geological Survey (Report OF 00-0325): 199-201.

Davis, C. 2000. “The Devil is in the Details (But How Do You Get Them in the Model?” *AAPG Bulletin* 84, No.6: 864.

Davis, J. C., and U. C. Herzfeld. 1993. *Computers in Geology: 25 Years of Progress*. New York: Oxford University Press.

Davis, J. C., and R. J. Sampson. 1992. “Trend-Surface Analysis: Early Computer Technique Remains Popular, Useful.” *Geobyte* (August 1992): 38-43.

Dawson, M., and M. Lim. 1993. “The Use of Information Systems and Technology in Petroleum Exploration.” *Geoscientist* 3 No. 2: 17-19.

Dengo, C. A. 1998. “Managing Technology in Today's Exploration Environment.” *AAPG Bulletin* 82, No.7:1442.

Derksen, C. R. M, and B.E. Haner. 1997. “Access and Cost of Geoscience Information in the Technology Era.” *Abstracts with Programs - Geological Society of America* 29, No.6: 173.

Dorn, G. A. 2002. “Computing and visualization.” *The Leading Edge* 21, No.6: 581-582, 584-586.

Dorn, G. A., K. Touysinhthiphonexay, J. Bradley, and A. Jamieson. 2001. “Immersive 3-D Visualization Applied to Drilling Planning.” *The Leading Edge* 20, No.12: 1389-1392.

Dunn, D. A. and T. Feldman . 1995. “GeoClio: Computer Bulletin Board and World-Wide-Web Site for the Study of the History of the Geosciences.” *Abstracts with Programs - Geological Society of America* 27, No.6: 255.

Egan, S. S., S. Kane, T.S. Buddin, G.D. Williams, and D. Hodgetts. 1999. “Computer Modelling and Visualisation of the Structural Deformation Caused by Movement Along Geological Faults.” *Computers & Geosciences* 25, no.3: 283-297.

Falconer, R. K. H., B. J. Scott, M. J. Kozuch, D.M. Johnston, T.H. Webb, K.R. Gledhill, and M.P. Chadwick. 1998. “GeoHazard Monitoring Networks: What Practical Uses They Bring to Hazard and Risk Management.” In Johnston, D M and P.A. Kingsbury, eds. *Proceedings of the Natural Hazards Management Workshop*. New Zealand: Institute of Geological & Nuclear Sciences Information Series (Report 45):11-17.

Ferderer, D. 2001. “A Data Management Life-Cycle.” Reston, VA: U. S. Geological Survey (Fact sheet FS 0163-00): 2.

Flynn, J. J. 1990. “3-D Computing Geoscience Update: Hardware Advances Set the Pace for Software Developers.” *Geobyte* (February): 33-35.

Ford, C E. 1998. “Supporting Fieldwork Using the Internet.” *Computers & Geosciences* 24, No.7: 649-651.

Fried, C. C., and J. E. Leonard. 1990. “Petroleum 3-D Models Come in Many Flavors.” *Geobyte* (Feburary): 27-30.

Fryer, S., J. Gregson, T. Connors, A.R. Poole, and B. Heise. 2001. “The National Park Service Digital Geologic Map Model: Transformation from Paper to Digital.” In Soller, D.R. ed., *Digital Mapping Techniques '01:Workshop Proceedings.* Reston: VA: U. S. Geological Survey (Report OF 01-0223): 87-100.

Giles, J.R.A, and K.A. Bain. 1995. “The Nature of Data on a Geological Map.” In Giles, J.R.A. ed. *Geological Data Management.* London: The Geological Society (Geological Society Special Publications 97): 33-40.

Giles, J. R. A. 1995. “The What, Why, When, How, Where and Who of Geological Data Management.” In Giles, J.R.A. ed. *Geological Data Management.* London: The Geological Society (Geological Society Special Publications 97): 1-45.

Gordin, D. N., and D.C. Edelson. 1997. “Using Visualization to Study Global Warming.” *Abstracts with Programs - Geological Society of America* 29, No.6: 180.

Gould, G. C., L.F. Marcus, A.P. Gueziec, G.F. Daleo, and D.P. Reddy. 1996. “Medical CT Scan Visualization and its Application for the Study of fossils Using Data Explorer and InterFormat.” *Journal of Vertebrate Paleontology* 16, No.3, Suppl.: 38.

Grimm, E. C. and J. Keltner. 1998. “The global pollen database.” *Palynology* 22: 242-243. Available at <http://www.ngdc.noaa.gov/paleo/pollen.html>.

Grossman, M. 1992. “Modeling Reality.” *IEEE Spectrum* (September): 56-60.

Gulick, V C, D.G. Deardorff, G.A. Briggs, T.A. Sandstrom, and Y. Hung. 2001. “A Virtual Collaborative Web Environment for Mars Surface Science Studies.” *Abstracts of the Annual planetary geologic mappers meeting.* Reston, VA: U. S. Geological Survey (Report OF 01-0018): 2.

Habermann, R.E., and D.M. Anderson. 2001. “What is the Role of the National Geophysical Data Center in Geoinformatics?” *Geological Society of America, 2001 Annual Meeting -- Abstracts with Programs* 33, No.6:175.

Hallmark, J. 1998. “Geoscientists' Access and Retrieval of Journal Articles in an Electronic World.” *Geological Society of America, 1998 annual meeting. Abstracts with Programs* 30, No.7: 189.

Harding, C., B. Loftin, and A. Anderson. 2000. “Visualization and Modeling of Geoscientific Data on the Interactive Workbench.” *The Leading Edge* 19, No.5: 506-511.

Hasenmueller, W. A. 1998. “Building Digital Geologic maps with a Contouring Program.” In Soller, David, ed., *Digital Mapping Techniques '98:Workshop Proceedings.* Reston, VA: U.S. Geological Survey (Report OF 98-0487): 67-71.

Henley, S. 1995. “Project Databases: Standards and Security.” In Giles, J.R.A. ed. *Geological Data Management.* London: The Geological Society (Geological Society Special Publications 97): 97-101.

Heron, D. W. 1994. “Geological hazards and resources in a GIS.” *ITMG Conference Papers.* Masterton, New Zealand: New Zealand Local Government Association, Information Technology Management Group.

Holland, S. M. and M.E. Patzkowsky. 1999. “Models for Simulating the Fossil Record.” *Geology* (Boulder) 27, No.6: 491-494.

Ingram, P. 1996. “Web developments and the Internet.” *Computers & Geosciences* 22, No.5: 579-584.

Iwao, Y., R. Izumiyama, A. Saitoh, N. Gurung, and T. Tsunetomi. 1998. “A Study on Global Earth Information System (GEIS)” in *Proceedings - International Congress of the International Association for Engineering Geology and the Environment* 8, No. 1: 585-590.

Jackson, I., K. Asch. 2002. “The Status of Digital Geological Mapping in Europe; the Results of a Census of the Digital Mapping Coverage, Approaches and Standards of 29 European Geological Survey Organisations in the Year 2000. *Computers & Geosciences* 28, No.6: 783-788.

Johnson, D. R., M. Ruzek, and M. Kalb. 1999. “Earth System Science; an Interdisciplinary Approach.” *GSA Today* 9, No. 8: 8-9.

Johnson, D.R; M. Ruzek, and M. Kalb. 2000. “Earth system science and the Internet.” *Computers & Geosciences* 26, No. 6: 669-676.

Jones, T. A. 1988. “Modeling Geology in Three Dimensions” *Geobyte* (February 1988): 14-20.

Jones, T.A. and J.E. Leonard. “Why 3-D Modeling” *Geobyte* (February 1990): 25-26.

Kacewicz, M. 1997. “3-D computer modeling of petroleum systems.” *Annual Meeting Abstracts - American Association of Petroleum Geologists and Society of Economic Paleontologists and Mineralogists* 6: 57.

Kafatos, E., T. El-Ghazawi, X.S. Wang, and R. Yang. 1999. “Earth Observing Data Systems in the Internet Era.” *Photogrammetric Engineering and Remote Sensing* 65, No.5: 540-541, 543, 545-548, 562-563.

Kaiser, J. 2000 “Fossil Databases Move to the Web.” *Science* 289, No. 5488 (29 Sep 2000): 2307.

Kingston, J. and S. Simson. 1997. “Building Corporate Knowledge Through Integrated Data Management.” *AAPG Bulletin* 81, No.8: 1388.

Korsell, K. 1997. “How an Innovative Consortium Shares Data to Compete on Analysis, not Data Acquisition.” *JPT (Journal of Petroleum Technology)* 49, No.1:49-50.

Kramer, J. H. 1997. “Future: geologic mapping as a digital art.” *Geological Society of America, 1997 annual meeting. Abstracts with Programs* 29, No.6: 472.

Krumbein, W.C. 1962. “The Computer in Geology.” *Science* 136,No. 3522, (June 29, 1962): 1087-1092.

Last, W. M . 1995. “Internet Listservers of Interest to Paleolimnologists.” *Journal of Paleolimnology* 13, No.3: 295-296.

Lehnert, K, Y. Su, C. H. Langmuir, B. Sarbas, U. Nohl. 2000. “A Global Geochemical Database Structure for Rocks.” *Geochemistry, Geophysics, Geosystems - G 3* 1 (24 May 2000): 16. (paper number 1999GC000026) Available at <http://g-cubed.org>.

Lennie, A. R. 2000. “Internet-Based Visualisation of Mineral Structures”. *Mineralogical Society Bulletin* 126 (Apr 2000): 3-5.

Lhotak, M., M.C. Boulter. 1995. “Towards the Creation of an International Database of Palaeontology.” *Geological Data Management.”* Geological Society Special Publications 97: 55-64.

Liro, L., K. Cline, and M. Kadri. 2001. “A New Paradigm for Data Integration and Collaboration Using 3-D Visualization Technology.” *The Leading Edge* 20, No.12: 1372-1376.

Loudon, T.V. 2000. “Geoscience after IT: Part B. Benefits for Geoscience from Information Technology, and an Example from Geological Mapping of the Need for a Broad View.” *Computers & Geosciences* 26 (Supplement): A5-A13.

Lulla, K., C. Evans, D.L. Amsbury, J. Wilkinson, K. Willis, J. Caruana, C. O'Neill, S. Runco, D. McLaughlin, M. Gaunce, M.F. McKay, and M. Trenchard. 1996. “The NASA Space Shuttle Earth Observations Photography Database; an Underutilized Resource for Global Environmental Geosciences.” *Environmental Geosciences* 3, No.1: 40-44.

Mancini, E. A., H. Chen, and D. J. Benson. 1998. “Geologic and Computer Modeling of Upper Jurassic Smackover Reef and Carbonate Shoal Lithofacies, Eastern Gulf Coastal Plain.” *AAPG Bulletin* 82, No.9: 1786-1787.

Merefield, J., D. Roche. 1993. “Geoenvironmental Information Technology: A Regional Group Initiative.” *Geoscientist* 3, No.3:25-26.

Miller, R.G. and J.W. Gardner. 1995. “Geoscience Data Value, Cost and Management in the Oil Industry.” In Giles, J.R.A. ed., *Geological Data Management.* London: The Geological Society (Geological Society Special Publications 97): 127-135.

Mitchell, K.J.R., A.M.K. Fritz, and R. Waldkirch. 1999. “Evolution of Digital Geologic Mapping at the North Dakota Geological Survey” In Soller, David R., ed., *Digital Mapping Techniques '99: Workshop Proceedings.* Reston, VA: U. S. Geological Survey (Report OF 99-0386): 177-181. , Available at <http://pubs.usgs.gov/openfile/of99-386/>

Morgan, K.M., A.B. Busbey III., R.N.Donovan, G.Walton, and N. Trenc. 1997. “Geologic Applications of Space Shuttle Imagery.” *Annual Meeting Abstracts - American Association of Petroleum Geologists and Society of Economic Paleontologists and Mineralogists* 6: 85.

NASA Advisory Council, Earth System Sciences Committee. 1986. *Earth System Science Overview: A Program for Global Change.* Washington, DC. 48 pp., May 1986.

National Research Council, Commission on Geosciences, Environment and Resources, Water Science and Technology Board, Board on Earth Sciences and Resources. 2000. *Seeing into the Earth: Noninvasive Characterization of the Shallow Subsurface for Environmental and Engineering Applications.* Washington, DC: National Academy Press. 129 pp.

National Research Council, Committee on Global Change. 1999. *Global Environmental Change: Research Pathways for the Next Decade.* Washington: National Academy Press. 595 pp.

National Research Council. 1997. *Research Required to Support Comprehensive Nuclear Test Ban Treaty Monitoring.* Washington, DC: National Academy Press. 135 pp.

National Research Council. 1999. *Distributed Geolibraries; Spatial Information Resources.* Washington, DC: National Academy Press. 119 pp.

Nelson, H. R., Jr. 2000. “Immersive Visualization.” *The Leading Edge* 19, No.5: 505

Nolen-Hoeksema, R., and C. Hart. 1991. “Stanford Rock Physics Database” *Geobyte* (Feb 1991): 20-26.

Normile, D. 2002 “’Earth Simulator’ Puts Japan on the Cutting Edge.” *Science* 295 (March 1, 2002):1631-1633.

Oreskes, N., K. Shrader-Frechette, and K. Belitz. 1994. *Science* 263, No. 5147 (Feb 4, 1994): 641-646.

Oreskes, N. 2000. “Why Believe a Computer? Models, Measures, and Meaning in the Natural World.” In Schneiderman, Jill S. ed., *The Earth Around Us.* New York: W.H. Freeman: 70-82.

Pack, S. 2001. “The Three-Dimensional Geologic Model as an Access Portal.” In Soller, David R. ed., *Digital Mapping Techniques '01: Workshop Proceedings.* Reston VA: U.S. Geological Survey (Report OF 01-0223): 25-27. Available at: <http://pubs.usgs.gov/of/2001/of01-223/>.

Phillips, C., and D. Clarke. 1998. “3D Modeling/Visualization Guides Horizontal Well Program in Wilmington Field.” *Journal of Canadian Petroleum Technology* 37, No.10: 7-15.

Pittman, J.G., and S.G. Brown. 2001. “Internet-Based Collaboration in Paleontology.” *Journal of Vertebrate Paleontology* 21, No.3, Suppl.: 89.

Pristas, R.S., and G.C. Herman. 1997. “Overview of New Jersey Geological Survey Digital Data Methods.” *Proceedings of a Workshop on Digital Mapping Techniques; Methods for Geologic Map Data Capture, Management, and Publication.* Reston, Virginia: U. S. Geological Survey (Report OF 97-0269): 39-41.

Ramshaw, S. R. 1995. “Geoscience Listservers and Newsgroups.” *Computers & Geosciences* 21, No. 6: 787-790.

Rankey, E.C., and W. L. Watney. 1994. “Analysis of Stratigraphic/ Sedimentologic Computer Models.” *1994 International Association for Mathematical Geology Annual Conference;Papers and Extended Abstracts for Technical Programs* 39, No.1: 293-298.

Rasmussen, K. 1995. “An Overview of Database Analysis and Design for Geological Systems.” In Giles, J.R.A. ed. *Geological Data Management.* London: The Geological Society (Geological Society Special Publications 97): 5-11.

Redfern, S.A.T. 1995. “Advances in Computer Modeling of Mineral Properties.” *Mineralogical Magazine* 59, No.4: 585-587.

Reid, J.C., M.A.Medina, and M.W. Carter. 2000. “Digital Geologic Map Production at the North Carolina Geological Survey.” In Soller, David, ed., *Digital Mapping Techniques '00: Workshop Proceedings.* Reston, VA: U. S. Geological Survey (Report OF 00-0325):189-190.

Rhyne, T.M. 1997. “Going Virtual with Geographic Information and Scientific Visualization.” *Computers & Geosciences* 23, No.4: 489-491.

Richard, D. 2000. “Development of an Internet Atlas of Switzerland.” *Computers & Geosciences* 26, No.1: 45-50.

Romanowicz, B. and D. Giardini. 2001. “The Future of Permanent Seismic Networks” *Science* 293 (September 14, 2001): 2000-2001.

Ross, J. A. 1998. “Providing Spatial Data and GIS Applications via the Internet.” In Soller, David, ed., *Digital Mapping Techniques '98: Workshop Proceedings.* Reston, VA: U.S. Geological Survey (Report OF 98-0487): 85-90.

Saunders, M.R. J.A. Shields, and M.R. Taylor. 1995. “Improving the Value of Geological Data: A Standardized Data Model for Industry.” In Giles, J.R.A. ed., *Geological Data Management.* London: The Geological Society (Geological Society Special Publications 97): 41-53.

Scotese, C.R., W.J. Nokleberg, J.W.H. Monger, I.O. Norton, L.M. Parfenov, A.I. Khanchuk, T.K. Bundtzen, K.M. Dawson, R.A. Eremin, Y.F. Frolov, K. Fujita, N.A. Goryachev, A.I. Pozdeev, V.V. Ratkin, S.M. Rodinov, I.S. Rozenblum, D.W. Scholl, V.I. Shpikerman, A.A. Sidorov, and D.B. Stone. 2001. *Dynamic Computer Model for the Metallogenesis and Tectonics of the Circum-North Pacific.* Reston, VA: U. S. Geological Survey (Report OF 01-0261).

Seber, D. C.O. Sandvol, C. Brindisi, and M. Barazangi. 2001. "Building the Digital Earth." *Geological Society of America, 2001 Annual Meeting -- Abstracts with Programs* 33, No.6: 176.

Seber, D., D. Steer, E. Sandvol, C. Sandvol, C. Brindisi, and M. Barazangi. 2000. "Design and Development of Information Systems for the Geosciences: An Application to the Middle East." *GeoArabia* (Manama) 5, No.2: 269-296.

Smith, T. M., T.R. Karl, and R.W. Reynolds. (2002) "How Accurate Are Climate Simulations? *Science* 296 (April 19, 2002): 483-484.

Smith-Rouch, L.S; K.E.Meisling, P.E. Hennings, and J.M. Armentrout. 1997. "Tectono-Stratigraphic Computer Experiments." *Annual Meeting Abstracts - American Association of Petroleum Geologists and Society of Economic Paleontologists and Mineralogists* 6: 108.

Soller, D. R. 1997. "The National Geologic Map Database: Support for Analytical Uses of Earth-Science Information." *Geological Society of America, 1997 annual meeting. Abstracts with Programs* 29, No.6: 472.

Soller, D.R. and T.M. Berg. 2001. "The National Geologic Map Database: A Progress Report." In Soller, D.R. ed., *Digital Mapping Techniques '01: Workshop Proceedings.* Reston VA: U.S. Geological Survey (Report OF 01-0223). Available at: <http://pubs.usgs.gov/of/2001/of01-223/>.

Soller, D.R. ed. 1999. *Digital Mapping Techniques '99 – Workshop Proceedings. Reston, VA:* U.S. Geological Survey (Report OF 99-386). Available at <http://pubs.usgs.gov/openfile/of99-386/>.

Soller, D.R. ed. 2000 *Digital Mapping Techniques '00 -- Workshop Proceedings.* May 17-20, 2000. Reston, VA: U.S. Geological Survey (Report OF 00-325). Available at <http://pubs.usgs.gov/openfile/of00-325/>.

Soller, D.R., R.Wahl, J.Weisenfluh, B. Brodaric, J. Hastings, R. Laudati, and R. Fredericks. 2001. "Progress Report on the National Geologic Map Database, Phase 3 - An Online Database of Map Information." Denver, CO: U. S. Geological Survey. Available at http://pubs.usgs.gov/of/2001/of01-223/contents.html.

Soller, D.R, and T.M. Berg. 2001. “The National Geologic Map Database: A Progress Report.” *Digital Mapping Techniques '01: Workshop Proceedings.* Reston, VA: U.S. Geological Survey (Report OF 01-0223): 51-57. Available at <http://pubs.usgs.gov/of2001/of01-223/>.

Souto Filho, J.d.D. 1998. “Numerical Modeling of Secondary Oil Migration: Definition of Oil Pathways in Laboratory Scale.” *AAPG Bulletin* 82, No.10: 1969.

Sowder, K.H., R.T. Hill, and P.N. Irwin. 2000. “The Process of Presenting GIS Information: Making GIS User-Friendly.” In Soller, D.R. ed., *Digital Mapping Techniques '00: Workshop Proceedings.* Reston, VA: U. S. Geological Survey (Report OF 00-0325): 173-174, 2000. Available at <http://pubs.usgs.gov/openfile/of00-325/>.

Stanford, L.R. and J.S. Freed. 1997. “Digital Map Production at the Idaho Geological Survey.” *Proceedings of a Workshop on Digital Mapping Techniques: Methods for Geologic Map Data Capture, Management, and Publication.* Reston, VA: U. S. Geological Survey (Report OF 97-0269): 31-34.

Staudigel, H., F. Albarede, D.L. Anderson, L. Derry, B. McDonough, H.F. Shaw, W.M. White, and A. Zindler. 2001. “Electronic Data Publication in Geochemistry: A Plea for ‘Full Disclosure.’” *Geochemistry, Geophysics, Geosystems - G 3,* 2001 (Paper 2001GC000234, 11 Oct 2001.)

Staudigel, H., F. Albarede, J. Blichert-Toft, J. Edmond, B. McDonough, S.B. Jacobsen, R. Keeling, C.H. Langmuir, R.L. Nielsen, T. Plank, R. Rudnick, H. F. Shaw, S.B. Shirey, J. Veizer, and W.White. 1998. “Geochemical Earth Reference Model (GERM): Description of the Initiative.” *Chemical Geology* 145, No.3-4: 153-159.

Stoss, F.W. 1998. “World Data Centers: Gateways to Geoscience Data.” *Geotimes* 43, No.10: 24-28.

Swinden, S. 2001. “Geoscience in the Dot-Com World: Changing with the Times.” *Geoscience Canada* 28, No.3: 105-111.

Takeda, A. S. Cox, and A.J. Payne. 2002. “Parallel Numerical Modeling of the Antarctic Ice Sheet.” *Computers & Geosciences* 28, No.6: 723-734.

Thoen, B. (1995) “Internet Resources for the Geosciences, with an Emphasis on GIS and Mapping. *Computers and Geoscience* 21, No. 6: 779-786.

Toll, D.G. and A.J. Oliver. 1995. “Structuring Soil and Rock Descriptions for Storage in Geotechnical Databases.” In Giles, J.R.A. ed. *Geological Data Management.* London: The Geological Society (Geological Society Special Publications 97): 65-71.

Une, H. 2001. “Toward the next stage of the Global Mapping Project: Successful Completion of Phase 1 with Release of Global Map Version 1.0” *Bulletin of the Geographical Survey Institute* 47: 13-19.

Vining, M.R. 1998. “Computer Modeling of the Evolution of Fluvial Systems.” *Transactions - Gulf Coast Association of Geological Societies* 48: 467-474.

Wagner, F.J. 1992. “A Brief History of Computer Contouring” *Geobyte* (June 1992): 10-11.

Walter, W. R., and H.E. Hartse, eds. 2002. *Monitoring the Comprehensive Nuclear-Test-Ban Treaty: Seismic Event Discrimination and Identification.* Switzerland: Birkhäuser Verlag AG. 289 p.

Wardlaw, B.R, N.R. Stamm, and D.R. Soller. 2001. “The U.S. Geological Survey National Paleontological Database.*”* In Soller, D.R. ed., *Digital Mapping Techniques '01; Workshop Proceedings.* Reston VA: U.S. Geological Survey (Report OF 01-0223) (no page numbers in electronic version). Available at: << http://pubs.usgs.gov/of/2001/of01-223/wardlaw.html >>.

Wells, B. and P. Duller. 1996. “Improving the Quality of Our Information Resource.” *Annual Meeting Abstracts - American Association of Petroleum Geologists and Society of Economic Paleontologists and Mineralogists* 5: 148.

Wendt, G.A and K.J. Lanfear. 1999. “Getting it off the Loading Dock: How the U. S. Geological Survey is Putting Scientific Information into the Hands of Users.” *Proceedings; Specialty Conference; Science into Policy; Water in the Public Realm.* American Water Resources Association Technical Publication Series 99-2: 141-146.

White, M. 1998. “The Application of Geological Modeling Systems in Radioactive Waste Disposal Projects.” *Electronic Geology* 1: 1-13.

Wood, W.B. 1997. “A Jeffersonian Vision for Mapping the World.” *Issues in Science and Technology 14,* No.1: 81-86.

Zeitlin, M.J. 2001. “The Present and Future of 3-D Visualization.” *The Leading Edge* 20, No.12: 1370-1371.

Zeitlin, M. 1992. “Visualization Brings a New Dimension to Oil Exploration and Production.” *Geobyte* (June 1992): 36-39.

Zhao, Y., F. Huang, and J. Guo. 1999. “Desertification research based on digital Earth”. IEEE International Geoscience and Remote Sensing Symposium, 1999. Vol. 1, pp.646-648, 1999.

Zwolinski, Z. 1996. “Polish Geoscience Resources on the WWW.” *Computers & Geosciences* 22, No.7: 831-832.

Bioscience References

Allen, M. T., C.E. Myers and M.A. Gluck. 2001. "Parallel Neural Systems for Classical Conditioning: Support from Computational Modeling." *Integrative Physiological & Behavioral Science* 36. No. 1 (January-March, 2001): 36-61.

Arzberger, P. and T.A. Finholt. 2002. "Data and Collaboratories in the Biomedical Community: Report of a Panel of Experts Meeting Held September 16-18, 2002 in Ballston, VA." White Paper. Available at <http://nbcr.ucsd.edu/Collaboratories.htm>.

Attwood, T.K. and C.J. Miller. 2001. "Methodological Overview -- Which Craft is Best in Bioinformatics?" *Computers and Chemistry* 25: 329–339.

Augen, J. (2002) "Bioinformatics and Information Technology: Reshaping the Drug Discovery Process." *Drug Discovery Today* 7, No. 11 (Information Biotechnology Supplement): 339-240.

Balant, L.P. and M. Gex-Fabry. 2000. "Modeling During Drug Development." *European Journal of Pharmaceutics & Biopharmaceutics* 50, No.1 (July, 2000): 13-26.

Bandak, F.A. 1995. "On the Mechanics of Impact Neurotrauma: A Review and Critical Synthesis." *Journal of Neurotrauma* 12, No. 4: 635-649.

Barnes, M.R. 2002. "Psychiatric Genetics in Silico: Databases and Tools for Psychiatric Geneticists." *Psychiatric Genetics* 12, No. 2: 67-73. Available at <http://www.psychgenetics.com/ >.

Bassingthwaighte, J.B. 2001. "ABME: On-line Submission, On-line Review." *Annals of Biomedical Engineering* 29, No. 5: 359.

Bayat, A. 2002. "Science, Medicine, and the Future: Bioinformatics." *BMJ* 324, No. 7344: 1018-1022.

Bevan, M. 2002. "Genomics and Plant Cells: Application of Genomics Strategies to Arabidopsis Cell Biology." *Philosophical Transactions of the Royal Society of London B Biological Sciences* 357, No. 1422: 731-736. Available at <http://www.jstor.org/journals/02643960.html>.

Bisby, F.A. "Quiet Revolution: Biodiversity Informatics and the Internet." *Science* 289, No. 5488 (29 Sep 2000): 2309-2312.

Bookstein, F.L. 1996. "Biometrics, Biomathematics and the Morphometric Synthesis." *Bulletin of Mathematical Biology* 58, No. 2: 313-365.

Bowen J. P., P.S. Charifson, P.C. Fox, M. Kontoyianni, A.B. Miller, D. Schnur, E.L. Stewart, C. Van Dyke. "Computer-Assisted Molecular Modeling: Indispensable Tools for Molecular Pharmacology." *Journal of Clinical Pharmacology* 33, No. 12: 1149-1164.

Brinkley, J. F., S.W. Bradley, J.W. Sundsten, and C. Rosse. 1997. “The Digital Anatomist Information System and its Use in the Generation and Delivery of Web-Based Anatomy Atlases.” *Computers & Biomedical Research* 30, No. 6: 472-503.

Burrough, P A. 2001. “GIS and Geostatistics: Essential Partners for Spatial Analysis.” *Environmental & Ecological Statistics* 8, No. 4: 361-377. Available at <http://www.kluweronline.com/issn/1352-8505>.

Butina, D., M.D. Segall, and K. Frankcombe. 2002. “Predicting ADME Properties in Silico: Methods and Models.” *Drug Discovery Today* 7, No. 11 Suppl.: S83-S88. Available at <http://www.drugdiscoverytoday.com>.

Butler, D. 1999. "Computing 2010: From Black Holes to Biology" *Nature* 402*,* Supp. 2 (December 1999): C70.

Cabeza, M. and A. Moilanen. 2001. “Design of Reserve Networks and the Persistence of Biodiversity.” *Trends in Ecology & Evolution* 16, No. 5: 242-248.

Calabrese, E.J. and R.B. Blain. 1999. “The Single Exposure Carcinogen Database: Assessing the Circumstances Under Which a Single Exposure to a Carcinogen can Cause Cancer.” *Toxicological Sciences* 50, No. 2: 169-185.

Charbel, F., J. Shi, F. Quek, M. Zhao, and M. Misra. 1998. “Neurovascular Flow Simulation Review.” *Neurological Research* 20, No. 2: 107-115.

Chicurel, M. 2002. “Cell Migration Research is on the Move.” *Science* 295, No. 5555: 606-609.

Claustres, M., O. Horaitis, M.Vanevski, and R. Cotton. 2002. “Time for a Unified System of Mutation Description and Reporting: A Review of Locus-Specific Mutation Databases.” *Genome Research* 12, No. 5: 680-688. Available at <http://www.genome.org/>.

Claverie, J. 1998. “Computational Methods for Exon Detection.” Molecular Biotechnology 10 No. 1: 27-48.

Claverie, J. 1997. Computational Methods for the Identification of Genes in Vertebrate Genomes Sequences.” *Human Molecular Genetics* 6, No. 10: 1735-1744.

Covert, M.W., C.H. Schilling, I. Famili, J.S. Edwards, I.I. Goryanin, E.Selkov, and B.O. Palsson. 2001. “Metabolic Modeling of Microbial Strains in Silico. *Trends in Biochemical Sciences* 26, No. 3: 179-186.

Davies, E. K. and Graham, R.W. 2002. “The Potential of Internet Computing for Drug Discovery.” *Drug Discovery Today* 7, No. 11 Suppl.: S99-S103. Available at <http://www.drugdiscoverytoday.com>.

Dayhoff, J.E. and J.M. DeLeo. 2001. “Artificial Neural Networks: Opening the Black Box.” *Cancer* 91, No. 8 Supplement: 1615-1635.

De Jong, H. 2002. “Modeling and Simulation of Genetic Regulatory Systems: A Literature Review.” *Journal of Computational Biology* 9: 67-103. Available at <http://www.liebertpub.com/CMB>

Department of Energy (2001) *Genomes to Life: Accelerating Biological Discovery.* Washington. (DOE/SC-0036) Accessed on line <http://www.doegenomestolife.org/roadmap/GTLcomplete_web.pdf>

Duckworth, D. M. and P. Sanseau. 2002. “In Silico Identification of Novel Therapeutic Targets.” *Drug Discovery Today* 7, No.11 Suppl.: S64-S69. Available at <http://www.drugdiscoverytoday.com>.

Edwards, J.L., M.A. Lane, and E.S. Nielsen. 2000. “Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop.” *Science* 289, No. 5488: 2312-2314.

Ekins, S. and S.A. Wrighton. 2001. “Application of In Silico Approaches to Predicting Drug-Drug Interactions.” *Journal of Pharmacological & Toxicological Methods* 45, No. 1: 65-69.

Endy, D. and R. Brent. 2001. “Modelling Cellular Behaviour.” *Nature* 409, No. 6818: 391-395.

Garcia, R.L.A. and S.P. Gutthann. 1998. “Use of the UK General Practice Research Database for Pharmacoepidemiology.” *British Journal of Clinical Pharmacology* 45, No. 5: 410-425.

Genoud, T., M.B.T. Santa Cruz, J.P. Metraux. 2001. “Numeric Simulation of Plant Signaling Networks.” *Plant Physiology* 126, No. 4: 1430-1437.

Glorioso, J. and N. Lemoine. 2002. “Electronic Submission, Review and Tracking System for Gene Therapy.” *Gene Therapy* 9, No. 1: 1. Available at <http://www.naturesj.com/gt/>.

Gram, L.F., J. Hallas, and M. Andersen. 2000. “Pharmacovigilance Based on Prescription Databases.” *Pharmacology & Toxicology* 86 (Supplement 1): 13-15.

Greenberg, S.A. “DNA Microarray Gene Expression Analysis Technology and its Application to Neurological Disorders.” *Neurology* 57, No. 5: 755-761.

Hokkanen, J.E.I. 1999. “Visual Simulations, Artificial Animals and Virtual Ecosystems.” *Journal of Experimental Biology* 202, No. 23: 3477-3484.

Holsheimer, J. 1998. “Computer Modelling of Spinal Cord Stimulation and its Contribution to Therapeutic Efficacy.” *Spinal Cord* 36, No. 8: 531-540.

Honig, B. 1999. “Protein Folding: From the Levinthal Paradox to Structure Prediction.” *Journal of Molecular Biology* 293, No. 2: 283-293.

Huelsenbeck, J.P. and B. Rannala. 1997. “Phylogenetic Methods Come of Age: Testing Hypotheses in an Evolutionary Context.” *Science* 276, No. 5310:7. 227-232.

Huret, J.L., P. Dessen, S. Le Minor, and A. Bernheim. 2000. “The ‘Atlas of Genetics and Cytogenetics in Oncology and Haematology’ on the Internet and a Review on Infant Leukemias.” *Cancer Genetics & Cytogenetics* 120, No. 2:000. 155-159.

Jang, S.H., M. G. Wientjes M, L.S. Au Jessie. 2001. “Determinants of Paclitaxel Uptake, Accumulation and Retention in Solid Tumors.” *Investigational New Drugs* 19, No. 2: 113-123.

Johnson B.L. 1995. “Applying Computer Simulation Models as Learning Tools in Fishery Management.” *North American Journal of Fisheries Management* 15, No. 4: 736-747.

Johnson, D.E., and H.I.W. Grushenka. 2000. “Predicting Human Safety: Screening and Computational Approaches.” *Drug Discovery Today* 5, No. 10: 445-454.

Kaminski, N. 2000. “Bioinformatics: A User's Perspective.” *American Journal of Respiratory Cell & Molecular Biology* 23, No. 6: 705-711.

Kao, C.M. 1999. “Functional Genomic Technologies: Creating New Paradigms for Fundamental and Applied Biology.” *Biotechnology Progress* 15, No. 3: 304-311.

Kaufman, M.H., R.M. Brune, D.R. Davidson, R.A. Baldock. 1998. “Computer-Generated Three-Dimensional Reconstructions of Serially Sectioned Mouse Embryos.” *Journal of Anatomy* 193, No. 3: 323-336.

Kelly, D. E., D.C. Lamb, S.L. Kelly. 2001. “Genome-Wide Generation of Yeast Gene Deletion Strains.” *Comparative & Functional Genomics* 2, No. 4: 236-242.

Klebe, G. 2000. “Recent Developments in Structure-Based Drug Design.” *Journal of Molecular Medicine* 78, No. 5: 269-281.

Kramer, J.A. 2001. “OmigaTM: A PC-Based Sequence Analysis Tool.” *Molecular Biotechnology* 19, No. 1: 97-106.

Kusnirikova, P. and E. Cellarova. 2000. “Sequence Databases for Molecular Biologists.” *Biologia* (Bratislava) 55, No. 6: 617-626.

Landers, E.S. R. Landgridge, and D.M. Saccocio. 1991. “Computing in Molecular Biology: Mapping and Interpreting Biological Information.” *IEEE Computer* 24, No. 11: 6-13.

Little, T J. 2002. “The Evolutionary Significance of Parasitism: Do Parasite-Driven Genetic Dynamics Occur Ex Silico?” *Journal of Evolutionary Biology* 15, No. 1: 1-9. Available at <http://link.springer.de/link/service/journals/00036/index.htm>.

Loehle, C., and D. Leblanc. 1996. “Model-Based Assessments of Climate Change Effects on Forests: A Critical Review.” *Ecological Modelling* 90, No. 1: 1-31.

Mack, R. and M. Hehenberger. 2002. “Text-Based Knowledge Discovery: Search and Mining of Life-Sciences Documents.” *Drug Discovery Today* 7, No. 11 Suppl.: S89-S98. Available at: <http://www.drugdiscoverytoday.com>.

McKenzie, F E. 2000. “Why Model Malaria?” *Parasitology Today* 16, No. 12: 511-516.

Mehr, R. 2001. “Modeling the Meta-Dynamics of Lymphocyte Repertoires.” *Archivum Immunologiae et Therapiae Experimentalis* 49, No. 2: 111-120.

Merrill, S.J. 1998. “Computational Models in Immunological Methods: An Historical Review.” *Journal of Immunological Methods* 216, No. 1-2: 69-92.

Miller, W. 2001. “Comparison of Genomic DNA Sequences: Solved and Unsolved Problems.” *Bioinformatics* 17 No. 5: 391-397.

Moir, W. H. and H.T. Mowrer. 1995. “Unsustainability.” Forest Ecology & Management 73, No. 1-3: 239-248.

Mouritsen, O. G., B. Dammann, H.C. Fogedby, J.H. Ipsen, C. Jeppesen, K. Jorgensen, J. Risbo; M. C. Sabra; M. M. Sperotto; M. J. Zuckermann. “The Computer as a Laboratory for the Physical Chemistry of Membranes.” *Biophysical Chemistry* 55, No. 995: 55-68.

Myers, E.W, G.G. Sutton, A.L. Delcher, I.M. Dew, D.P. Fasulo, M.J. Flanigan, S.A. Kravitz, C.M. Mobarry, K.H.J. Reinert, K.A. Remington, E.L. Anson, R.A. Bolanos, H.H. Chou, C.M. Jordan, A.L. Halpern, S. Lonardi, E.M. Beasley, R.C. Brandon, L. Chen, P.J. Dunn, Z. Lai, Y. Liang, D.R. Nusskern, M. Zhan, Q. Zhang, X. Zheng, G.M. Rubin, M.D. Adams, C.J. Venter. 2000. “A Whole-Genome Assembly of Drosophila.” *Science* 287, No. 5461: 2196-2204.

Nadkarni N.M., and G.G. Parker. 1994. “A Profile of Forest Canopy Science and Scientists: Who We Are, What We Want to Know, and Obstacles We Face: Results of an International Survey.” *Selbyana* 15, No. 2: 38-50.

Naisberg, Y. 2001. “Biophysical Vision Model and Learning Paradigms About Vision: Review.” *Medical Hypotheses* 57, No. 4: 409-418.

National Academy of Sciences. 1970. *The Life Sciences: Recent Progress and Application to Human Affairs. The World of Biological Research, Requirements for the Future.* Washington, DC: National Academy Press.

National Institutes of Health. 2002. "Complex Biological Systems Initiatives: NIGMS Initiatives to Promote Interdisciplinary Approaches to Research on Complex Biomedical Problems" Accessed December 8, 2002. <http://www.nigms.nih.gov/funding/complex_summary.html>

National Institutes of Health. 1998. “Report of the NIH Model Organism Database Workshop December 7-8, 1998.” Accessed online December 8, 2002. <http://www.nhlbi.nih.gov/meetings/modeldb/report.htm>.

Nazmul, K.M., T.Yoshida, S.L. Rivera, V.M. Saucedo, B.E. G.S. Oh. “Global and Local Neural Network Models in Biotechnology: Application to Different Cultivation Processes.” *Journal of Fermentation & Bioengineering* 83, No. 1: 1-11.

New Scientist. 1999. "The Heart that Just Won't Die." *New Scientist* (20 March 1999) accessed online December 7, 2002 <http://www.newscientist.com/hottopics/ai/heart.jsp>

Oh, J.M.C., F. Brichory, E. Puravs, R. Kuick, C.Wood, J.M. Rouillard, J. Tra, S. Kardia, D. Beer, and S. Hanash. 2001. “A Database of Protein Expression in Lung Cancer.” *Proteomics* 1, No. 10: 1303-1319.

Orr, L., Govindjee (no second name provided). 2001. “Photosynthesis and the Web.” *Photosynthesis Research* 68, No. 1: 1-28.

Palmer, A.J., A. Brandt, V. Gozzoli, C. Weiss, H. Stock and H. Wenzel. 2000. “Outline of a Diabetes Disease Management Model: Principles and Applications.” *Diabetes Research & Clinical Practice* 50 (Supplement 3): S47-S56.

Perez, V.J.L., and P.L. Carlen. 2000. “Gap Junctions, Synchrony and Seizures.” *Trends in Neurosciences* 23, No. 2: 68-74.

Pieper U., N. Eswar, A.C. Stuart, V.A. Ilyin, and A. Sali. MODBASE, a database of annotated comparative protein structure models. Nucl. Acids Res. 30, 255-259, 2002.

Pitt, D.G., and D.P. Kreutzweiser. 1998. “Applications of Computer-Intensive Statistical Methods to Environmental Research.” *Ecotoxicology & Environmental Safety* 39, No. 2: 78-97.

Polacco, M., E. Coe, Z. Fang, D. Hancock, H. Sanchez-Villeda, and S. Schroeder. “ MaizeDB: A Functional Genomics Perspective.” *Comparative & Functional Genomics* 3, No. 2: 128-131.

Poulin, M., M. Coste, F. Straub, and L. Ector. 2001. “Diatoms and Databases: A Short Review.” *Vie et Milieu* 51, No. 1-2: 29-35.

Prusinkiewicz, P. 1998. “Modeling of Spatial Structure and Development of Plants: A Review.” *Scientia Horticulturae* 74. No. 1-2: 113-149.

Radford, S.E., and C.M. Dobson. 1999. “From Computer Simulations to Human Disease: Emerging Themes in Protein Folding.” *Cell*, 97, No. 3: 291-298.

Railsback, S.F. 2001. “Concepts from Complex Adaptive Systems as a Framework for Individual-Based Modelling.” *Ecological Modelling* 139, No. 1: 47-62.

Rawlings, C.J., J.P. Fox. 1994. “Artificial Intelligence in Molecular Biology: A Review and Assessment.” *Philosophical Transactions of the Royal Society of London B Biological Sciences* 344, No. 1310: 353-363.

Robert, C., C. Guilpin, and A. Limoge. 1998. “Review of Neural Network Applications in Sleep Research.” *Journal of Neuroscience Methods* 79, No. 2: 187-193.

Rodrigues, A. D., G.A. Winchell, and M.R. Dobrinska. 2001. “Use of In Vitro Drug Metabolism Data to Evaluate Metabolic Drug-Drug Interactions in Man: The need for Quantitative Databases.” *Journal of Clinical Pharmacology* 41, No. 4: 368-373.

Rogers, R. and P.S. Johnson. 1998. “Approaches to Modeling Natural Regeneration in Oak-Dominated Forests.” *Forest Ecology & Management* 106, No. 1: 45-54.

Rooney, K.F., E. Snoeck and P. Watson. 2001. “Modelling and Simulation in Clinical Drug Development.” *Drug Discovery Today* 6, No. 15: 802-806.

Roth, A., V. Mosbrugger, and A. Wunderlin. 1998. “Computer Simulations as a Tool for Understanding the Evolution of Water Transport Systems in Land Plants: A Review and New Data.” *Review of Palaeobotany & Palynology* 102, No. 1-2: 79-99.

Rust, A. G., E. Mongin, and E. Birney. 2002. “Genome Annotation Techniques: New Approaches and Challenges.” *Drug Discovery Today* 7, No. 11 Suppl.: S70-S76. Available at <http://www.drugdiscoverytoday.com>.

Sabri, S., F. Richelme, A. Pierres, A.M. Benoliel, and P. Bongrand. 1997. “Interest of Image Processing in Cell Biology and Immunology.” *Journal of Immunological Methods* 208, No. 1: 1-27.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. "DNA Sequencing with Chain-Terminating Inhibitors," *Proc. Acad. Natl. Sci. U.S.A.* 74: 5463-5467.

Sarwal, A. and A.P. Dhawan. 2001. “Three Dimensional Reconstruction of Coronary Arteries from Two Views.” *Computer Methods & Programs in Biomedicine* 65, No. 1: 25-43.

Schaefer, C., L. Grouse; K. Buetow; R.L. Strausberg. “A New Cancer Genome Anatomy Project Web Resource for the Community.” *Cancer Journal* 7, No. 1: 52-60.

Schilling, C.H., S. Schuster, B.O. Palsson, and R. Heinrich. 1999. "Metabolic Pathway Analysis: Basic Concepts and Scientific Applications in the Post-Genomic Era." *Biotechnology Progress* 15, No. 3: 296-303.

Schmajuk, N.A., L. Cox, and J.A. Gray. 2001. "Nucleus Accumbens, Entorhinal Cortex and Latent Inhibition: A Neural Network Model." *Behavioural Brain Research.* 118, No. 2: 123-141.

Schoonbaert, D. 2000. "Tropical Medicine, the Internet and Current Trends in Biomedical Communication." *Annals of Tropical Medicine & Parasitology* 94, No. 7: 661-674.

Scriver, C.R. and P.M. Nowacki. 1999. "Genomics, Mutations and the Internet: The Naming and Use of Parts." *Journal of Inherited Metabolic Disease* 22, No.4: 519-530.

Sigmund, K. 1998. "Complex Adaptive Systems and the Evolution of Reciprocation." *Ecosystems* 1, No. 5: 444-448.

Sipper, M. 1998. "Fifty Years of Research on Self-Replication: An Overview." *Artificial Life* 4, No. 3: 237-257.

Sophia T., and C.A. Ouzounis. 2000. "Recent Developments and Future Directions in Computational Genomics." *FEBS Letters* 480: 42-48. (Federation of European Biochemical Societies.)

Spafford E.H. 1994. "Computer Viruses as Artificial Life." *Artificial Life* 1, No. 3: 249-265.

Sun, E., and F.E. Cohen. 1993. "Computer-Assisted Drug Discovery-A Review." *Gene* (Amsterdam) 137, No. 1: 127-132.

Sun, W., and P. Lal. 2002. "Recent Development on Computer Aided Tissue Engineering: A Review." *Computer Methods & Programs in Biomedicine* 67 (February): 85-103. Available at <http://www.elsevier.com/locate/cmpb>.

Suvarna, S.K., and M.A. Ansary. 2001. "Histopathology and the 'Third Great Lie'. When Is an Image not a Scientifically Authentic Image?" *Histopathology* (Oxford), 39, No. 5: 441-446.

Tatusov, R.L., E.V. Koonin, D.J., and Lipman D.J. 1997. "A Genomic Perspective on Protein Families." *Science* 278, No. 5338:631-637.

Taylor, T. and C. Massey. 2001. "Recent Developments in the Evolution of Morphologies and Controllers for Physically Simulated Creatures." *Artificial Life* 7, No. 1: 77-87.

Van Helden, J., A. Naim, R. Mancuso, M. Eldridge, L. Wernisch, D. Gilbert, and S.J. Wodak. "Representing and Analysing Molecular and Cellular Function Using the Computer." *Biological Chemistry* 381, No. 9-10: 921-935.

Van Pee, M., and D. Berckmans. 1999. "Quality of Modelling Plant Responses for Environment Control Purposes." *Computers & Electronics in Agriculture* 22, No. 2-3: 209-219. B

Venter, J.C., M.D. Adams, G.G. Sutton, A.R. Kerlavage, H.O. Smith, and M. Hunkapiller. 1998. "Shotgun Sequencing of the Human Genome," *Science* 280: 1540-1542.

Wang, J.Z. 2001. "Wavelets and Imaging Informatics: A Review of the Literature." *Journal of Biomedical Informatics* 34, No. 2: 129-141.

Webb, B. 2000. "What Does Robotics Offer Animal Behaviour?" *Animal Behaviour* 60, No. 5: 545-558.

Weinstein, J.N. 1998. "Fishing expeditions," *Science* 282: 627-628.

Weinstein, J.N. 2001. "Searching for Pharmacogenomic Markers: The Synergy Between Omic and Hypothesis-Driven Research." *Disease Markers* 17, No. 2: 77-88.

Weir, M., M. Swindells, and J. Overington. 2001. "Insights Into Protein Function Through Large-Scale Computational Analysis of Sequence and Structure." *Trends in Biotechnology* 19, No.10 Supplement: S61-S66.

Welch, T.J. 2002. "Bioinformatics: The Role and Limitations of Patents." *Drug Discovery Today* 7, No. 11 Suppl.: S77-S82. Available at <http://www.drugdiscoverytoday.com>.

Whiten, S., S.D. Smart, J.C. McLachlan, J.F. Aiton. 1998. "Computer-Aided Interactive Three-Dimensional Reconstruction of the Embryonic Human Heart." *Journal of Anatomy* 193, No. 3: 337-345.

Wishart, D.S. and S. Fortin. 2001. "The BioTools Suite: A Comprehensive Suite of Platform-Independent Bioinformatics Tools." *Molecular Biotechnology* 19, No. 1: 59-77.

Working Group on Biomedical Computing, Advisory Committee to the Director of the National Institutes of Health. 1999. "The Biomedical Information Science and Technology Initiative." Betheseda, Maryland: National Institutes of Health.

Yao, T. 2002. "Bioinformatics for the Genomic Sciences and Towards Systems Biology: Japanese Activities in the Post-Genome Era." *Progress in Biophysics & Molecular Biology* 80: 23–42.

Zeng, J. 2000. “Mini-Review: Computational Structure-Based Design of Inhibitors that Target Protein Surfaces.” *Combinatorial Chemistry & High Throughput Screening* 3, No. 5: 355-362.

Ziegler, C., W. Goepel, H. Haemmerle, H. Hatt, G. Jung, L. Laxhuber, H.L. Schmidt, S. Schuetz, F. Voegtle, A. Zell. 1998. “Bioelectronic Noses: A Status Report. Part II.” *Biosensors & Bioelectronics* 13, No. 5: 539-571.