Volume 4, Issue 9, September 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Some studies on Databases Relevant to Biotechnology

N. Deepak Kumar*
Department of CSE & SV University
India

Dr. A. Ramamohan Reddy
Department of CSE & SV University
India

Abstract— The development and the use of databases become increasingly important to biotechnology in research and development, in industry, and in the public eye. The particular importance of biotechnology information is especially obvious in connection with the necessity of molecular biology databases. This can be seen from the high value of information in the frame of genome projects, as the U.S. Human Genome Program and the U.S. Plant Genome Research Program, which include large-scale projects combining mapping and sequencing with data collection and distribution.

The "information highway" exists per se in form of a highly developed information infrastructure, and the political changes brought the advantage that we have no unbridgeable frontiers. The task of biotechnology is to draw a map, not only in genome research, but also in the field of biotechnology information and bioinformatics with the aim of targeting the information highway. The combination of biotechnology and information technology is a challenge for both fields and should be a strategy for the future.

Keywords— Bioinformatics,molecular,genome,gene,protein,DBMS.

I. INTRODUCTION

The availability of high-quality, up-to-date and comprehensive information is an important requirement in biotechnology and its applications in the ever widening fields of medicine, pharmacy, agriculture, food industry and the environmental sciences. Advances in biotechnology, especially in genome research, depend to an increasing extent on which required information is made available and how it is used. The growing amount of data, especially the genome projects, deliver an enormous number of sequence data, meaning that collecting, processing and disseminating these data is only possible with the help of modern information technology and international co-operation[1].

Modern biotechnology is highly information-dependent and uses a wide variety of information sources and information technologies[2]. The consequences of rapid developments in biotechnology with its tremendous volume of data and in informatics with its potential for processing and using date are:

1. In relation to research –
- the creation of a new scientific field: Bioinformatics

2. In relation to infrastructures –
- very large databases and smaller, more highly specialized databases
- highly sophisticated software for processing and using these databases
- efficient communication networks for access to databases and information exchange
- establishment of information centers for collection, processing and distribution of information
- comprehensive information services

Many databases are available in various types of media: Online via a number of networks and hosts, or on CD-ROM, diskette, magnetic tape, or as a printed version. The overall growth in the online database industry during the past year can be traced through the statistics: 300 online databases were registered in the 1979 edition of *Directory of Online Database[5]*. In the 1993 edition, the Gale Directory of Databases registered profiles more than 5200 online databases, among them about 250 with relevance to biotechnology, and more than 3200 database products offered in portable form, among them about 150 with relevance to biotechnology (MAR-CACCIO, 1993). The number of records stored in these databases increased from 52 million (1975) to about 5 billion (1993)[5,7].

II. METHODOLOGY

The compilation of databases in the fields of biotechnology (CRAFTS-LIGHTY, 1986; POETZSCH, 1986, 1988; ALSTON and COOMBS, 1992) and the investigations of the very different information needs of users reflecting the multidisciplinary character of biotechnology bear witness to the diversity of the necessary information and of the wide variety of offered databases. Derived from the results of these investigations, we can state that the following types of information are necessary for biotechnology:

Factual information. Especially sequence information (nucleic acid sequences, protein sequences) is of greatest importance to biotechnology, but also map information, structure information, and property information. Commercial and financial information, especially for the biotechnology industry, play an increasing role.

Bibliographic information. In addition to literature information, patent information is important to biotechnology.

Referral information (Directory information), including information on research institutions, Research & Development projects companies, products, culture collections.

Full-text information with the complete text from journals, newsletters, biotechnology and related regulations, encyclopedias, market research reports.

Databases relevant to biotechnology can be classified

- according to subject areas:
  The subject area is usually the determining point for user selection of a database. In this connection, an important factor influencing biotechnology information is the interdisciplinary of biotechnology with various sciences, application areas and related fields.
- according to the type of stored information:
  Factual databases, bibliographic databases, referral databases, full-text databases.

2.1 Factual Databases

The close link between biotechnology and information technology is particularly evident in relation to nucleic acid and protein acid sequences stored in factual databases. In the literature, author report that the DNA sequence database will become as important as the Periodic Table of Elements. Factual databases are most important to biotechnology because these databases serve not only as an information tool but as a direct research tool. Factual databases provide a research instrument which exists at the interface between subject area and information technology, whereby the scientist increasingly assumes the role of producer and user of information, and bear witness to the increasing influence of information technology on the research process.

Depending on the type of stored information, the most important factual databases in biotechnology are in the following:

Nucleotide sequences:
GenBank (USA)
EMBL Data Library (EC, UK)
DNA Data Bank of Japan (Japan)
GENESEQ-Patents Sequence Database (UK)
REGISTRY file (USA)
MEDLINE Molecular Sequence Data (USA)
RNA Data Bank (Germany)
Vector Bank (USA)
dbEST (USA)

Protein sequences:
PIR Protein Sequence Database (USA, in collaboration with Germany and Japan)
SWISS-PORT Protein Sequence Database (Switzerland)
REGISTRY file (USA)
GENESEQ-Patents Sequence Database (UK)
GenPept (USA)
PseqIP (France)

Species-specific mapping data:
Genome Data Base (GDB, USA)
Online Mendelian Inheritance in Man (OMIM, USA)
Genomic Database of the Mouse (GBASE, USA)
The Encyclopedia of the Mouse Genome (USA)
Escherichia coli Genetic Stock Center (CGSC, USA)
Fly Base (USA)
EcoMap (USA)
ACEDB (a Caenorhabditis elegans database, UK)
AAtDB(Arabidopsis thaliana database, USA)
Plant Genome Database (PGB, USA)

Structures:
Protein Data Bank (USA)
Cambridge Structural Database (UK)
CarbBank (Carbohydrate Structure Database, USA)

BEILSTEIN (Germany)
REGISTRY file (USA)
CASREACT (USA)
ChemInform RX(Germany)

Restriction enzymes:
REBASE (Restriction Enzyme Databank, USA)

Enzymes:
BRENDA (B Raunschweiger Enzyme Database, Germany)
DBEMP (DataBase on Enzymes and Metabolic Pathways, Russia)

Microorganisms/culture collections:
DSM Catalogue (Germany)
MSDN Central Directory (UK)
MINE (EC, Germany)
ATCC Catalogues (USA)
MiCIS (UK)

Cell cultures/hybridomas:
Immunoclone Database (France in collaboration with Germany and UK)
INTERLAB Network: Cell Line Data Base, B Line Data Base, Molecular Probe Data Base, ImmunoClone Data Base, Genotoxicity and Carcinogenicity Database (all Italy)

In the following, the most important sequence databases (GenBank, EMBL Data Library, PIR International) are described in more detail.

GenBank (Genetic Sequences Databank)
Producer: National Center for Biotechnology Information, USA
Contents: GenBank contains DNA and RNA sequences, bibliographic citations, related information such as sequence descriptions, source organisms, sequence lengths etc., and software packages for using GenBank.
Access: Online, e.g., via STN International, GENIUSnet, BIONET Online Service and FTP access; CD-ROM; Magnetic Tape
Cooperation: EMBL European Molecular Biology Laboratory, DDBJ DNA Database of Japan
Information Sources: Scientific Journals, Direct Data Submission
Additional Service; E-mail Service, NCBI Data Repository, NCBI Newsletter

EMBL (EBI) Data Library
Producer: European Molecular Biology Laboratory, Heidelberg, Germany (until August, 1994); European Bioinformatics Institute, Hinxton Park, Cambridge, UK
Products and Services available from the EMBL (EBI) Data Library:

- Databases on CD-ROM: EMBL NUCLEOTIDE SEQUENCE DATABASE, SWISS-PROT PROTEIN SEQUENCE DATABASE and 31 related databases: PROSITE (Protein pattern database), ENZYME (Database of EC nomenclature), ECP (E. coil map database), EPD (Eukaryotic promoter database), REBASE (Restriction enzyme database), FlyBase (Drosophila genetic map database), TFD (Transcription factor database), TRNA (tRNA sequences, RRNA (Small subunit rRNA sequences), BERLIN (5S sequences), KABAT (Proteins of immunological interest) etc., and software for search and retrieval of data (EMBL Search, CD-SEQ)
- Network File Servers using Electronic Mail, FTP, Gopher server
- European Molecular Biology Network (EMBnet): The main activity of EMBnet is the daily distribution of all new sequence data via the computer network to 16 nationally mandated nodes.
- Sequence Searching Services: BLITZ, Mail-Quick search and Mail-Fast A are services that allow external users to search the DNA and protein sequence databases via electronic mail.
- Sequence Data Submission: AUTHORIN SOFWARE PACKAGE, Submission Form (Computer-readable copies or printed copies), Data Submission by electronic mail or by post.

PIR International Protein Sequence Database
Producers: Protein information Resource (PIR) at the National Biomedical Research Foundation(NBRF), USA; Martinsried Institute for Protein Sequences (MIPS) at the MAX Planck Institute for Biochemistry, Germany; Japan International Protein Information Database (JIPID) at the Science University of Tokyo, Japan
Contents: The PIR contains descriptions of partial and whole protein sequences including function of protein, taxonomy, sequence features of biological interest, how sequence was experimentally determined, unambiguously determined residues within the sequence, and citations to relevant literature.
Data Input: PIR (USA) provides approx. 50% of data; MIPS (Germany) provides approx. 35% of data; JIPID (Japan) provides approx. 15% of data.

Access: CD-ROM (‘Atlas of Protein and Genomic Sequences”) Magnetic Tape Online via EMBnet, BIONET and through other networks on a number of file servers

PIR international includes the following databases:

PIR1 (annotated and classified entries), PIR2 (preliminary entries), PIR3 (unverified entries), PATCHX (yet unprocessed by PIR), MIPSH (Yeast Protein Sequences), ECON (E. coli data set), Alignment Database, NRL-3D Sequence-Structure Database, and software for processing sequence data

Source: Scientific Journals, Direct Data Submission, EMBL Data Library

2.2 Bibliographic Databases

Bibliographic databases contain citations, mostly with abstracts, to the published literature, i.e., journal articles, patents, reports, dissertations, conference proceedings, books, etc. There are about 100 bibliographic databases relevant to biotechnology. Especially the access to patent information is of great importance to everyone working in the field of biotechnology. Patents have been called the lifeblood of the biotechnology industry.

Major bibliographic databases in biotechnology and related fields area:

BIOSIS Previews, CAS ONLINE, DERWENT Biotechnology Abstracts, Chemical Engineering and Biotechnology Abstracts, CSA Life Sciences Collection, BioBusiness, BioCommerce Abstracts and Directory, PASCAL: Biotechnologies, Biotechnology Citation Index, BioExpress, Current Awareness in Biological Sciences and others, together with

- **Application-related databases in**
  Medicine and Pharmacy: MEDLINE, EMBASE, CANCERLIT, AID Database, AIDSLINE, International Pharmaceutical Abstracts, DERWENT Drug File, Pharmline, Bioethicsline
  Agriculture and Nutrition: CAB ABSTRACTS, AGRIS, AGRICOLA, AgBiotech News and information, DERWENT Crop Protection File, Food Science and Technology Abstract, Foods adlibra
  Environment: PLLUAB, ULIDAT, ENVIROLINE, TOXLINE, AQUASCI
  Chemistry: Chemical Abstracts, Chemical Business NewsBase, Chemical Industry Notes, Analytical Abstracts
  Engineering: COMPENDEX, INSPEC, VtB verfahrenstechnische Berichte
- **multidisciplinary databases relevant to biotechnology:**
  Science Citation Index, Current Contents, and others
- **patents databases relevant to biotechnology:**
  DERWENT Biotechnology Abstract, DERWENT World Patents Index, PATDPA, PATOSEP, PATOSWO, INPADOC, Current Patents, PATFULL, JAPIO, CLAIMS, Drug Patents International

2.3 Referral Databases

Referral databases (Directory databases) contain information on research activities, Research & Development projects, institution and company profiles, products, and services. Especially in connection with the necessity of business information in biotechnology, those database containing company information are of great importance. There are about 50 referral databases with relevance to biotechnology and a number of printed directories which are also available as databases.

- **Biotechnology:** WHO-WHAT-WHERE in Biosciences and Biotechnology (WWW/BIKE, Germany; printed versions: Biotechnology Das Jahr- und Adreβbuch(POETZSCH, 1993) and Biotechnology Directory Eastern Europe (LŰCKE and POETZSCH, 1993), BIOREP (BIOtechnology Research Projects, EC), BioCommerce Abstracts and Directory (UK; printed version: The U.K. Biotechnology Handbook), BEST Biotech (USA), BIOTEC (Cuba), Leading Biotechnology Companies (USA), Federal Bio-Technology Transfer Directory (USA)
- **Multidisciplinary** (with information on biotechnology): CORDIS (EC), Corporate Technology Database (USA), ICC Directory (U.K), Japanese Corporate Directory (Japan), Directory of French Companies (French), Who Supplies What? (Germany), American Business Directory (USA), Research Centers and Services Directory (USA), Who’s Who in Technology? (USA)
- **Pharmacy:** Pharmaprojects (UK), Pharmacontacts (UK), AIDSDRUGS (USA)
- **Agriculture:** AGREP (EC), CRIS/USDA (USA), TEKTRAN (USA), European Directory of Agrochemical Products (UK)
- **Environment:** UFORDAT (Germany), DETEQ (Germany), DEQUIP (Germany)
- **Information sources:** Directory for Biotechnology Information Resources (USA), Listing of Molecular Biology Databases (USA), Information Sources in Biotechnology (Germany), I’M Guide (EC), CUADRA/GALE’s Database Directory (USA)

2.4 Full-Text Databases

Full-text databases contain the complete text of original publications, e.g., journal articles, newsletters, newspapers, regulatory documents, encyclopedia, market research reports.

- **Databases containing texts of journals and newsletters,** e.g., European Biotechnology Information Service, Biotech Knowledge Sources, Biotechnology Investment Opportunities, Biotech Business, Biotechnology Newswatch, BioWorld Online (USA), Japan Report: Biotechnology, Genetic Technology News, Applied Genetics

News, ADIS DrugNews, Drug Information Full-text, MEDTEXT, Pharmaceutical and Healthcare Industry News, AIDS Newsletter

- **Databases containing (complete) texts of statutes and legislations,** e.g., DIOGENES, Federal Register, Biological Monitoring Database
- **Databases containing (complete) texts of encyclopedias,** e.g., IMSWorld New Product Launch Letter, KIRK-OTHMER Online, The MERCK INDEX Online
- **Databases containing (complete) texts of market research reports,** e.g., MARKET-FULL, MarkIntel, KOBRA

III. RESULT

The use of databases and other information sources as an aid to increase the public perception of biotechnology: The lack of public acceptance of biotechnology is also a result of the lack of information, since there is a direct link between information and attitude. Researchers in the public perception of biotechnology agree that attempts to improve access to scientific information are highly desirable (GRINDLEY and BENNETT, 1993). Specific measures should be taken to enhance public perception mostly through the availability of objective information, especially in connection with biotechnology's impact on human health through the development of new pharmaceutical, diagnostic, and other medical products.

VI. CONCLUSION

The complexity of data and databases connected with the necessity to crosslink different (types of) information which are part of different (types of) databases, to combine different forms of information representation, to extend the numeric data by means of supplementary, descriptive information, to use standardized or easily translatable formats which must be interconnected in order to integrate individual databases in a global concept. The integration of databases from various producers and structures in systems which have a single, unified administration and allow a homogeneous access to the various heterogeneous data present. For bibliographic databases, the Commission of the European Communities recommends the creation of a "Common core Database" by the bunching of central biotechnology databases for the prevention of duplicates, overlapping, etc.

REFERENCES

[1] ALSTON, y., COOMBS, j. (1992), Biosciences, Information Sources and Services, New York: Stockton Press.
[2] CRAFTS-LIGHTLY, A. (1986), Information Sources in Biotechnology, Weinheim: VCH Verlagsgesell-schaft.
[3] GRINDLEY, J.N., BENNETT, D.J. (1993), Public perception and the socio-economic integration of biotechnology, in: Biotechnologia 20, 89-102.
[4] LŰCKE, E.-M., POETZSCH, E. (1993), Biotechnology Directory Eastern Europe, Berlin-New York: de Gruyter.
[5] MARCACCIO, K. Y. (1993), Gale Directory of Databases, Vol. 1: Online Database, Detroit: Gale Research Inc.
[6] MEWES, H.-W. (1990), Workshop Computer Applications in Biosciences, Book of Abstracts, p. 11, Martinsried.
[7] POETZSCH, E. (1986), Faktographische informationsfonds zur Biotechnologie, Berlin: WIZ.
[8] POETZSCH. E. (1993), Bio Technologie Das Jahrund Adreβbuch 93/94, Berlin: polycom Verlags-gesellschaft.

