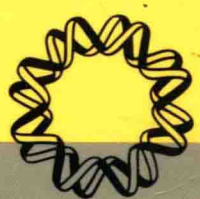
THE BRITISH LIBRARY

SCIENCE TECHNOLOGY
AND INDUSTRY

INTRODUCTION TO BIOTECHNOLOGY INFORMATION



SCIENCE AND TECHNOLOGY

INTRODUCTION TO BIOTECHNOLOGY INFORMATION

Written by the staff of
The British Library
Biotechnology Information
Service

Edited by M EUESDEN

THE BRITISH LIBRARY

Published by
The British Library
Science Reference and Information Service
25 Southampton Buildings
London
WC2A 1AW

For further information on SRIS publications,
contact Paul Wilson in the Publications Department,
tel: 071 323 7472

Printed in Great Britain by
Antony Rowe Ltd, Chippenham, Wiltshire

British Library Cataloguing in Publication Data
Introduction to biotechnology information.
1. Biotechnology. Information sources
I. British Library and Environmental Sciences Group
II. Science Reference and Information Service
660.6072

ISBN 0-7123-0762-1

ACKNOWLEDGEMENTS

A number of past and present members of the British Library's Biotechnology Information Service contributed directly or indirectly to this book. These people have drawn on their knowledge of biotechnology information to make significant contributions to the text. The chapter on official publications is largely due to text prepared by Lesley Downs; information on online sources and access was provided by Rupert Lee; Sunil Nandi prepared text for areas of the book dealing with scientific information and patents; and Martin Pinnell was responsible for the final chapter on bioethics.

CONTENTS

1 SCIENTIFIC INFORMATION

Introduction

Many new techniques have emerged from scientific laboratories over the last two decades to revolutionise industry through biotechnological advances. The advances stem from the breakthroughs in molecular biology which have revealed the secret workings of DNA. The codes which guide protein synthesis in cells can now be manipulated to make useful products that are beyond the capabilities of traditional industrial chemistry. The understanding of these molecular processes is the prelude to using living systems to make industrially useful products. Science leads and technology follows. Scientific sources of information are therefore valuable for understanding the products of today and anticipating the products of tomorrow.

One of the most important applications of biotechnology is in the pharmaceutical area, for the generation of highly specific chemical complexes to mimic or control human metabolism, and to aid the immune system to fight disease.

Some notable areas where scientific discoveries have made an impact, or are expected to make an impact, on industry are:

Genetic fingerprinting. This is now successfully marketed by ICI through Cellmark Diagnostics and used around the world to identify individuals on the basis of the banding patterns made by the electrophoretic separation of nucleic acids from a sample of body tissue. It improves on the traditional fingerprinting technique because it can also show relatedness of individuals, and the information can be digitised for easier storage.

Biocontrol using genetically engineered microorganisms. It is possible, for example, to spray onto crops certain bacteria which generate chemicals toxic to crop pests, but which do not pollute the soil with harmful residues. Other variations include the generation of a symbiotic relationship to provide the host plant with protection or nutrition in the form of nitrogen fixed from the atmosphere.

Transgenic organisms. Animals and plants can have genetic material from a foreign organism inserted into their genome to confer attributes of use to investigators. Two notable cases of this are the production of a transgenic mouse containing oncogenes which make it more susceptible to carcinogens and thus a suitable experimental subject for the testing of potentially carcinogenic substances. The transgenic tomato has been given

improved characteristics which can be achieved in a much smaller time than the usual route of propagation.

Biosensors. The highly specific binding characteristics of monoclonal antibodies and enzymes can be harnessed to an electrical or chemical indicator system to give sensitive measurements of important chemicals, for example blood sugar.

BIBLIOGRAPHIES OF BIOTECHNOLOGY INFORMATION

The most comprehensive guide to all kinds of biotechnology information, including science, is:

Information Sources in Biotechnology, by A Crafts-Lighty. Macmillan, 1986. 2nd. edition.
It covers the practical use of sources of information in biotechnology, and contains extensive bibliographies.

Also:

Information Sources in the Life Sciences, by H V Wyatt. Butterworths, 1987. This covers articles on biotechnology, microbiology, databanks and genetics.

Biotechnology Information '86, edited by R Wakeford. IRL Press, 1987. This book contains the proceedings of a conference held in 1986 and organised by The British Library and IRL Press. It reviews and discusses the technology of data collection, the organisation and running of databanks and the practical experience of people working in commercial intelligence departments.

Primary sources

There are a number of comprehensive reference works dealing with science and technology, but the field is moving so fast that the information tends to date quite quickly. Important reference works include:

Comprehensive biotechnology and bioengineering, edited by Murray Moo-Young [et al]. Blackwell, 1985. 4 volumes.

Biotechnology: a comprehensive treatise in 8 volumes, edited by H-J Rehm and G Reed. Weinheim, VCH, 1981-87.

Biotechnology in agriculture and forestry, edited by Y P S Bajaj. Springer-Verlag, 1986- . 10 volumes.

Good introductory texts are:

Biotechnology, by J E Smith. London, Arnold, 1988. (Studies in Biology Series, Institute of Biology).

Biotechnology: principles and applications, by I J Higgins [et al]. Blackwell, 1985.

Secondary sources

TITLE LISTS

Current Contents is published by the Institute of Scientific Information in Philadelphia, USA. It is a family of titles which each publish the contents pages of around 100 top journals in a particular field. Three titles are particularly relevant to biotechnology:

Agriculture, Biology and Environmental Science
Life Sciences
Clinical Science

The contents pages, arranged by subject, are published within weeks of appearance of the original issue. Articles of interest can be found by scanning all the relevant journal titles. The authors' addresses are given in the back to enable the reader to send reprint requests. Articles may also be obtained through the publisher of *Current Contents* in the form of 'original article tear sheets'. Contact: Institute for Scientific Information, 3501 Market Street, Philadelphia, PA 19104, USA.

Sheffield University Biomedical Information Service (SUBIS) publish a large number of current awareness bulletins in specific fields such as monoclonal antibodies, plant biotechnology and recombinant DNA. They indicate titles and sources of new articles. The time lag is greater than that for *Current Contents*, but coverage is greater as 1500 journals are scanned. The entries are in classified subject order and forthcoming events are also included. The bulletins are also available on computer diskettes. Contact: SUBIS, Sheffield University, Sheffield S10 2TN, UK.

The British Library Document Supply Centre (DSC) offers a range of services including a special subscription service to supply the contents pages of named journals, as they are received by The Library. This gives the user access to titles not covered in Current Contents. It is fast and DSC is also a reliable source from which to obtain copies of articles of interest. For more information, see Appendix H.

SCIENCE CITATION INDEX

Produced by the Institute for Scientific Information, Philadelphia, the *Science Citation Index* is unique in that it serves not only as a retrieval tool but can also be used for bibliographic studies of the scientific literature.

Articles may be retrieved by author, title word (in two-word permutations), corporate body or by cited article. The latter is a useful means of indexing a 'concept' which might otherwise be difficult to search on. Essentially, the researcher searches for the names of authors who are prominent in a particular field and whose articles are cited by others. If a paper is cited by an author, then the author of the cited paper is likely to be involved in related work. There are annual and five year cumulations which makes retrospective searches more convenient to carry out. They

are much used by scientific authors who avidly scour the pages to find if anyone has cited their work!

The files are available online as SCISEARCH, and are also available on compact disc.

The ISI *Atlas of science: biotechnology and molecular genetics* 1987 uses bibliographic analysis of the SCI database to group papers which tend to be co-cited (ie there is a significant number of papers in which any two other given papers are both cited). This gives a measure of the relatedness of two named papers.

ABSTRACTS

There are numerous abstracting and indexing services in the biological and chemical sciences, from *Chemical Abstracts* to *Current Biotechnology Abstracts*. The general abstracts are expensive and occupy a lot of shelf space, so are practical only for larger libraries. The subject-based services are more targeted but might not cover a specific aspect of biotechnology in what is obviously a very broad field.

There is inevitably a delay between publication of the original and publication of the abstract (typically two to six months) so although the entries are often conveniently laid out in classified subject order they are probably best suited to retrospective searching rather than current awareness. Equivalent online files are available as an alternative in many cases and their extensive indexing and speed do provide a useful current awareness service. It is expensive to carry out an online search but the overall costs and benefits may compare well with the hardcopy version over a period of time, depending on what type of use is made of the system. Briefly, for a small number of searches on a wide variety of databases it is likely to be cheaper to search online than take out numerous subscriptions to abstracting services of which 90% of the material may not be of use over the course of the year.

Some of the important sources are listed below.

General abstracts

Biological Abstracts has a KWIC (keyword in context) subject index, author index and taxonomic index. Its coverage is restricted to the biological sciences.

Chemical Abstracts is strong in biotechnology and has detailed indexes for subject words, chemical substances, formulae and authors. Where necessary these are cross-indexed. The abstracts in the life sciences section are published every two weeks and there are five year cumulations which are useful for retrospective searching. There is good coverage of patents and also reviews, reports, books and conference proceedings. Sections of interest include Fermentation and bioindustrial chemistry, Waste treatment and disposal.

Index Medicus and *Excerpta Medica* (various) are medically oriented but useful for pharmaceuticals. Both use controlled indexing terms to aid retrieval.

The Commonwealth Agricultural Bureaux publish abstracts collections in various fields eg *Animal Breeding Abstracts and International Biodeterioration.* They are monthly and indexed by author and subject.

Abstracts in the field of biotechnology

These cover selected journal titles and the selection may not include subjects of interest to the user so it is necessary to check their scope. Some information about these titles is given below, indicating frequency, price, coverage and size.

Biotechnologies, Centre national de la recherche scientifique, Paris.
10 issues per year; 1,580 FF; 700 citations/issue.

Main sections: Legislation; applied research; methods, procedures, technologies; industrial applications, economical aspects.

Indexes: Author, subject.

Abstracts, where given, are in French. Non-English titles have an English translation.

Derwent Biotechnology Abstracts, Derwent, London.
Twenty-six issues per year; £500; Coverage 1,300 journals, plus conference proceedings, meetings reports, patents (basics, not equivalents, although US equivalents were covered up to 1988); 500 citations/issue.

Main classes: Range from engineering to cell culture.

Indexes: Author, corporate body, subject.

Also covered are worldwide patents including European and PCT patents. All abstracts are in English irrespective of the original language and there are diagrams to supplement the text where necessary. Not included in the coverage are products of biotechnology in non-industrial fields eg therapy. Brewing and fermented foods are also not covered. Main sections are microbiology, pharmaceuticals and biocatalysis (around 100 citations each per issue); others are engineering, chemistry, agriculture, food, energy, other chemicals, cell culture, biocatalysis, purification, waste disposal.

Biotechnology Research Abstracts, Cambridge Scientific, Cambridge, Mass.
Six issues per year; $535; Coverage 5000 journals; 800 citations per issue.

Main sections: Patents; genetic engineering; immobilisation; cell culture; products of biotechnology; fermentation and process engineering.

Indexes: Author, subject

Current Biotechnology Abstracts, Royal Society of Chemistry, Cambridge
Twelve issues per year; £299; Coverage 400 journals; European, PCT, UK and USA patents form around 20% of CBA's content; 350 citations/issue.

Main sections: Genetic manipulation; fermentation technology and pharmaceuticals (around 60 citations per issue). Also business; legal issues and safety; monoclonal antibodies; enzymology; single-cell protein; energy production; agriculture; chemical industry; food and other industrial areas.

Indexes: Subject; chemical name; companies but not authors.

Current Advances in Genetics and Molecular Biology,
Pergamon, Oxford.
Twelve issues per year; DM990.

Main sections: General and molecular genetics; prokaryotic genetics; viral genetics and eukaryotic genetics.

Indexes: Author index, no abstracts.

Derived from Current Awareness in Biological Sciences database (coverage 3,000 journals), available online through Orbit. Series includes titles in biochemistry and microbiology. Document delivery service available.

Telegen Reporter, Bowker A&I Publishing, New York.
No longer published; Coverage 400 journals; 275 citations/issue.

Main sections: Markets and issues; applications; research.

Indexes: Author; keyword; industrial area; source.

Has a definite bias toward industry and commerce. Includes short patent citations, from US Patent and Trademark Office, plus reviews.

Bioinvention, OMEC, Washington DC.
Twelve issues per year; $310; 100 citations/issue.

Abstracts of US biotechnology patents with the date of granting, application number and classmark.

COMPARISON WITH ONLINE SERVICES

Mention has been made of the relative merits of using hardcopy sources and online sources. It would seem that given computer access that this would always be preferable because of the ease of retrieval (virtually every word in the entry is indexed) and the breadth of databases available through the same terminal.

One of the main differences is cost. Computer processing time is very expensive, typically £2-£3 per minute, with additional charges for each reference which is 'downloaded'. If a searcher has access to a good library with many of the hardcopy versions of these databases, then they may well prefer to use these. Students researching a project are unlikely to spend a lot of money on online searching, but an industrial user may well regard the saving in time a worthwhile investment.

It is simpler and often cheaper (providing a person's time is not valued too highly) to browse through hardcopy sources. They are easier to read than a computer monitor and also do not pose the problems of time pressure which an online searcher is likely to experience. Because of the time delay between publication of an original article and its appearance in the abstract journal there is a problem in using hardcopy for current awareness. Online files are published electronically and are available as soon as the records have been updated. In some cases certain journals or subject areas are given priority which means that the online file may contain records referring to articles published only a few weeks earlier. Relevant titles can be picked up routinely by running a saved search activated automatically every month with results posted to the user.

It is unlikely that the equipment for online searching will pose much of a problem for potential online searchers as any microcomputer can be linked into data networks to carry out the interrogation of remote computers. Indeed a tiny microcomputer tucked in the corner of a room can give unparalleled access to scientific databases whereas hardcopy versions consume a great deal of valuable space. (For a more detailed discussion of online information retrieval see Chapter 3).

OTHER SECONDARY SERVICES

Research dissertations are valuable sources of information on experimental methods. They are not widely available for consultation but the following titles may be of interest, and would enable the user to apply for an interlibrary loan.

Dissertation Abstracts International, University Microfilms International, Ann Arbor, Michigan, USA.

Index to Theses Accepted for Higher Degrees in the Universities of Great Britain and Ireland. Aslib, London.

If dissertations are not available from one of the microfiche services there is always the possibility of obtaining an inter-library loan from the institution involved.

The following may be useful for biotechnology information in non-English language sources:

Russian material: This is well indexed in *Referativnyi Zhurnal* which has sections covering a number of subjects, some in English. The British Library Document Supply Centre produce a guide, *Journals in Translation*, which includes Soviet titles.

Japanese material: The best source of scientific information from Japan is the online database JICST, (Japanese Information Centre for Science and Technology). An English language version is available through the database host STN. Contact: JICST, 5-2 Nagatacho 2-chome, CPO Box 1478, Tokyo, Japan.

RESEARCH DIRECTORIES

There are a number of directories which give details of research projects in the field of biotechnology. The basic information given is the title of the research project, principal investigators and period of funding. In some cases there are also details of the size of the grant given.

Directory of Research in Biotechnology (SERC Biotechnology Directorate) gives detailed accounts of research projects in the public sector, funded by SERC, and supplies the names of the project organisers.

Biorep is a European directory of research in biotechnology, prepared by the Library and Documentation Centre of the Royal Netherlands Academy of Arts and Sciences (KNAW). It is available as an online database on the European Commission Host Organisation (ECHO). There is also a slightly abbreviated version available as hardcopy.

Biomed is a medical research directory also available from the European Commission.

Current Research in Britain (The British Library): Four parts, one covering biological sciences. Its scope is wider than that of the SERC publication.

Directory of British Biotechnology (Longmans/DTI) Briefly lists research fields of commercial and non-commercial organisations. It gives estimates of the range of funding on R&D and the number of personnel employed.

British Expertise in Science and Technology (Longman Cartermill): This is a database of research personnel, and is now available on compact disc, by subscription. Contact: Longman Cartermill Ltd, Technology Centre, St Andrews, Fife KY16 9EA, UK. Tel 0334 77660

Scientific data sources

Two especially important areas of biotechnology data provision concern culture collections and sequence databanks. Living cells constitute a production environment, something like a miniature factory, and nucleic

acid sequences are the assembly lines for making specific products. This is a simplification of what technologists actually do to harness living processes, but it does emphasise the importance of these once purely academic resources. Information about them is important for biotechnologists who are looking for successful expression systems to make new products.

CULTURE COLLECTIONS

Culture collections are repositories of organisms maintained in a suitable environment, well-characterised and authenticated. The host institutions provide catalogues of their holdings and also continue to optimise systems for propagation and storage of the cultures. In particular, they are likely to be rich sources of information of the metabolic characteristics of a number of organisms. With suitable search facilities it is possible to seek out an organism with desired properties.

A list of culture collections is given by V F McGowan and V B D Skerman in the *World directory of collections of cultures of microorganisms*. The 2nd edition, 1982, lists 356 culture collections in 56 countries.

In addition, culture collections offer identification services for newly isolated strains and may carry out contract research. Certain designated centres have taken on an important new role in the field of patenting in biotechnology - they act as International Depository Authorities (IDA's). Under the regulations of the Budapest Treaty 1977, patents citing a living organism as an important part of the invention being disclosed must deposit that organism with an IDA. The IDA will not release such deposits unless licensed to do so, or if required to supply the strain to an independent third party for the purposes of testing.

In the UK there are a number of major culture collections specialising in a certain taxonomic range, such as the National Collection for Yeast Cultures, and the European Centre for Animal Cell Cultures. Elsewhere there are major collections which are more centralised such as the ATCC, DSM and the Belgian Coordinated Collections of Microorganisms. (Details of these collections are given in Appendix A.)

Searching for culture information can present a problem because catalogues from individual collections present data in different ways. Now, however, there are a number of initiatives to promote the unification of microbial data recording. The DTI in the UK made efforts to remedy the particular problem of having national resources distributed among a number of major collections. For a period of time, staff of the Laboratory of the Government Chemist (LGC) Biotechnology Unit gathered data from the UK national collections and created a single database called the Microbial Culture Information System (MiCIS) - see below. This and other databases can be conveniently accessed through an international electronic network for microbiologists, the Microbial Strain Data Network (MSDN). (See overleaf for more details).

MICROBIAL CULTURE INFORMATION SERVICE

The database is now complete (although recent updates have not yet been made). The project has been discontinued by the DTI and the database has been passed to the ICECC in Braunschweig, home of the Deutsche Sammlung von Mikroorganisment (DSM).

The database contains minutely coded data about strains, their physiology, genetics, origins and so on. It is an online reference work in itself and can also be used to generate hardcopy catalogues.

It is possible to access the system directly, or via a gateway on MSDN. (See below).

MICROBIAL INFORMATION NETWORK FOR EUROPE (MINE)

This project was initially funded by the Commission of the European Communities (CEC) Biotechnology Action Programme. Its objective is to integrate information from all EEC culture collections. Following a review of the project, funding will continue from October 1990 and it is intended to start an online file in the latter half of 1991. There are at present nine European nodes supported by the European Commission, and Sweden will join as a non-EEC member. The aim of the project is to create a core data set from all contributing countries, using compatible formats, enabling anyone to locate useful strains at European centres. The enquirer then approaches the appropriate centre to obtain full strain data. Essentially MINE is providing catalogue data in contrast to databases such as MiCIS which provide full strain data. Ultimately, however, MiCIS will be absorbed into the MINE project, and it is already converted to MINE format. Contact: MINE Project Leader, International Mycological Institute, Ferry Lane, Kew, Surrey TW9 3AF, UK.

MICROBIAL STRAIN DATA NETWORK (MSDN)

This was set up as a result of a 1983 international workshop, under the auspices of CODATA, the World Federation for Culture Collections and the International Union of Microbiological Societies. It is host to a number of microbiological databases and provides gateways to others. Among the databases available are: Hybridoma Databank, Information Centre for European Culture Collections, National Collection for Yeast Cultures and the World Data Centre for Collections of Microorganisms. MSDN operates a central directory which guides the user to collections holding organisms with the characteristics they are looking for. It also links microbiologists around the world on an electronic mail network. The system is accessible through British Telecom's Telecom Gold. Contact: Elaine Ross, 307 Huntingdon Road, Cambridge CB3 0JX, UK. (See also Chapter 3, page 27).

MICROBIOLOGICAL RESOURCE CENTRES (MIRCEN)

This is a loose network of 15 centres around the world which coordinates the preservation of microbial gene pools. It also organises the arrangements for making the information available to developing countries. The project was set up by UNESCO. It publishes *MIRCEN News* and has a UK centre at the University of Kent.

DATABANKS

Databanks are centralised repositories of data. They are usually thought of as being a computerised resource although this is not necessarily the case. Their main application is in the storage of massive amounts of data which describe a particular system, and there are already large databank resources in the physical sciences

Biological science also makes use of databanks. *The Registry of Toxic Effects of Chemical Substances (RTECS)* has been produced since 1971 by the National Institute for Occupational Safety and Health (NIOSH) in the USA, and is used for chemical toxicology searching. It can be used for straightforward text searching of chemical substances but also offers the facility for chemical sub-structure searching - that is, it can retrieve data which would otherwise be inextricably buried in the text.

SEQUENCE DATABANKS

Since the elucidation of the genetic code molecular biologists have had increasing control over the means of gene expression, to the point now where a sequence of nucleic acids in a gene can be translated by a computer to reveal the equivalent protein product, and even predict some of its characteristics and properties. The synthesis of proteins to a high degree of specificity is of great potential use in the pharmaceutical and agricultural industries. The string of nucleotides making up a nucleic acid, or amino acids making up a protein, can now be rapidly analysed. The resulting sequence information may be reported in a scientific paper, but will also be deposited by the author with one of the major sequence databanks. When searching for sequence data it is no longer of use searching through bibliographic sources because there may be insufficient keys to identify the required sequence. Also, sequence data offers the possibility of performing useful computerised manipulations (such as computer translation, homology searching and projecting three dimensional structures).

The following are important resources:

GenBank

This is a major nucleic acid sequence databank produced at the Los Alamos National Laboratory, Los Alamos, NM 87545, USA. It now contains total sequence data exceeding 30 million base pairs in length, from a number of organisms.