Chapter 1

What is SciFinder® Scholar?

1.1 Introduction

Knowing how to search and use information is a key factor in successful teaching and research. However, the information world is so complex that staff and students are often overwhelmed. Further, quality information may be expensive and academic institutions may find it hard to provide even a basic level of service.

It has not always been like this. Perhaps, until the third quarter of the twentieth century, universities were able to give access to the major journals, indexing tools, and reference materials needed by their scholars. Much has changed with the onset of the electronic age, with declining budgets, with increases in the price and the number of journals, and with the development of new research areas.

One excellent solution is offered through SciFinder Scholar. Before understanding how to use SciFinder Scholar, it is necessary to understand what Scholar is and how it fits into the world of information retrieval.

1.2 The Primary and Secondary Literature

1.2.1 Primary Literature: Journals

The number of journals in the sciences alone presents a major problem. For example, one of the world's major science database producers, the Chemical Abstracts Service (CAS), covers around 8000 journals. For original articles with the publication year 1999, the CAS bibliographic database (CAPLUS) contains almost 900 000 records. Almost two-thirds of these articles are from journals produced by just over 3000 publishers, and approximately one-third of these records were published by scientific societies, academies, institutes, associations, and various university presses. This vast literature clearly presents problems not only for scientists who need to keep up with relevant literature in their field but also for librarians whose function is to serve their scientific community.

Much has been written on the subject of scientific publication, on the pressures on journal subscriptions, and on the influence of governments, capital markets, and technology. However, it should be remembered that scientists have benefited from both the society and the commercial publishers and are indeed fortunate to have such excellent primary and secondary sources of scientific information that are the envy of other disciplines (particularly the humanities and social sciences). Scientists' problems are not in the area of collection or indexing but rather in the size and cost of basic information needs. It helps if scientists and librarians work cooperatively with publishers, who are now responding much more favorably to the needs of institutions. For example, libraries are finding that substantial discounts are available through package deals with the major publishers.

While further changes in the scientific information industry will occur, what will always underpin the industry will be the need for the scientists to publish their discoveries and in turn for scientists to search for literature in their area.

Increasingly, publications will appear less in print and more directly through the Internet. Among the advantages of the latter is that it not only allows more direct access at the desktop but also allows linking of documents, for example, through citations. This worldwide Web of publications will grow, and as the Web search engines become more sophisticated, scientists will be able to search the vast primary literature more directly and conveniently.

1.2.2 Secondary Literature: Abstracting and Indexing Services

Indexing and abstracting services will continue to provide valuable services not only because they provide access to literature that predates the Web but also because they will continue to build databases that will be vital to the information needs of the scientists.

There are many databases that cover the sciences, and Table 1.1 summarizes some bibliographic databases that currently contain an excess of two million records. Whether the scientist should search these databases, or even more subject-specific and smaller databases, essentially depends on the type of research involved and to the cost and the availability of the service. If there

Table 1.1 Major scientific bibliographic databases and database producers. SciFinder Scholar contains CAPLUS and MEDLINE

Database producer	Database name[a]	Area covered	Number of records[b]
American Institute of Aeronautics & Astronautics	AEROSPACE	Aerospace research	2.2
National Agricultural Library	AGRICOLA	Agriculture	3.6
BIOSIS	BIOSIS®	Life sciences	12.7
CAB International	CABA	Agriculture and related sciences	3.9
CAS	CAPLUS	Multidisciplinary science	18[c]
Engineering Information Inc.	COMPENDEX	Engineering and technology	4.8
Elsevier Science	EMBASE	Medical and biological sciences	8.4
US Department of Energy	ENERGY	Energy	3.9
American Geological Institute	GEOREF	Geosciences	2.3
International Atomic Energy Agency	INIS	Nuclear research and technology	2
Institution of Electrical Engineers	INSPEC	Physics, electronics, and computing	6.6
US National Library of Medicine	MEDLINE	Medical sciences	11.4
Institute for Scientific Information	SCISEARCH®	Multidisciplinary science	18.8
US National Library of Medicine	TOXLINE®	Toxicology of drugs and chemicals	2.7
Derwent Information Ltd	WPINDEX	Patents	10

[a] On the STN network.
[b] Approximately (million) to end of 2000.
[c] However, a major backfile was added in 2001, and the current figure (September 2001) is 19.7 million.

is a single, inexpensive, and specific database that completely covers the information needs of the searcher, then the searcher indeed is fortunate.

However, for comprehensive retrieval, the scientist usually needs either to search a number of databases or to search databases that cover broad areas. The former option is most readily realized either through on-line or localized networks, and among the world's major international on-line networks in the sciences are Dialog *http://www.dialog.com*, Questel/Orbit *http://www.questel.orbit.com/english/index.htm*, and STN *http://www.cas.org/stn.html*. The effective use of on-line networks requires very substantial training, and access to these networks generally is through a pay-as-you-use cost structure, which is not necessarily ideal for institutions that need to provide access to information to a large number of users. Nevertheless, it helps if scientists are aware of the additional opportunities provided through these sources and work with their library staff when access through them may be of specific benefit.

In the sciences, four of the world's biggest databases (i.e. those with more than 10 million records in Table 1.1) are produced by BIOSIS *http://www.biosis.org/home_deluxe.html*, CAS *http://www.cas.org*, the National Library for Medicine (NLM) *http://www.nlm.nih.gov*, and the Institute for Scientific Information (ISI) *http://www.isinet.com*. Generally, scientists looking for broad coverage explore some of these databases first. It is important for scientists to know the areas that the different products cover, their specific features, and their various benefits if they are to perform efficient and effective searches.

Searching for information requires many skills and the process involves the application of the 'scientific method' in a manner similar to the conduct of a laboratory experiment. The analogy is summarized in Figure 1.1, and the suggestion is that university courses should include instructions on all

Step	Scientific method	Information retrieval
1	Knowledge of the subject	Knowledge of the coverage and content of databases, of potential alternative information resources, and of search options
2	Requirement for new knowledge	Requirement for information on a topic
3	Proposal for research or experiment	Conceptualization of initial search query
4	Conduct of experiment	Initial search
5	Observation and interpretation of results	Careful examination of initial answer sets
6	Revision or improvement of experimental design	Revision of search query based on observation of initial answers
7	Revised experiment	Exploration of alternative search options
8	Final outcome of experiment	Final answer set

Figure 1.1 The parallel between scientific method and information retrieval.

aspects of information retrieval and on the various products available, and instruction on the technical aspects of the scientific discipline! This will help scientists follow the 'scientific method' in information retrieval.

1.3 The SciFinder Scholar Option

SciFinder Scholar *http://www.cas.org/SCIFINDER/SCHOLAR* provides desktop access to several databases from CAS and to the bibliographic database from the US NLM (Table 1.2), and thus covers information from two of the world's largest database producers. The only other single source of all of these databases is the STN network.

The databases have very extensive coverage across the sciences. To illustrate this, Figure 1.2 shows a breakdown of records in CAPLUS (>30 000 records) and MEDLINE® (>14 000 records) for Schools and Departments in

Table 1.2 Databases in SciFinder Scholar

Database	Database name[a]	Coverage	Number of records[b]
CAS bibliographic database	CAPLUS	More than 8000 journals, and patents from more than 30 organizations from 1947[c,h]	>19 million
NLM bibliographic database	MEDLINE	More than 4300 journals from 1958[d]	>12 million
CAS chemical substance database	REGISTRY	All substances from 1967[c]	>33 million (substances)
CAS regulated chemicals database	CHEMLIST	Substances on national inventories and some US State lists[e]	>220 000
CAS chemical catalogs database	CHEMCATS	Substances in chemical catalogs from more than 600 suppliers[f]	>920 000 (substances)
CAS chemical reaction database	CASREACT	Mainly organic reactions from selected journals from 1974[g]	>4 million reactions

[a] With the exception of the MEDLINE file, all files are available on the STN network only.
[b] September 2001.
[c] Although many pre-1967 substances are also covered *http://www.cas.org/casdb.html*.
[d] *http://www.nlm.nih.gov/hinfo.html*.
[e] *http://www.cas.org/CASFILES/chemlist.html*.
[f] *http://www.cas.org/CASFILES/chemcats.html*.
[g] *http://www.cas.org/CASFILES/casreact.html*.
[h] Note added in proof. Backfile to 1907 was added in December 2001.

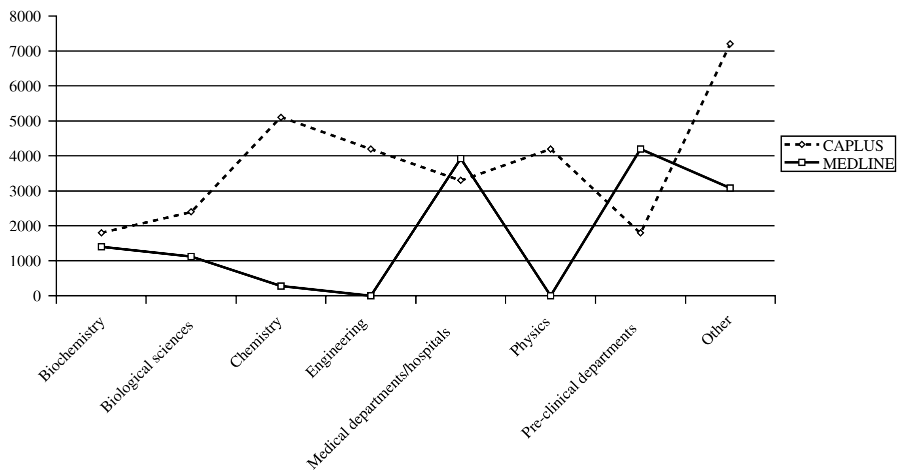


Figure 1.2 Number of records for discipline areas in Australian universities in CAPLUS (>30 000 records) and MEDLINE (>14 000 records) for 1995 to 1999. Data was obtained through searches in the Corporate Source Field in the CAPLUS and MEDLINE files on STN. Records from individual schools and departments were then combined into the broad categories illustrated. (Note: 'Other' in CAPLUS includes *inter alia* Environmental Sciences, Geosciences, Material Sciences, and entries in which only the University was listed. 'Other' in MEDLINE includes *inter alia* Dentistry, Health Sciences, Nursing, Psychology, and Veterinary Science.)

Australian universities from 1995 to 1999. While the breakdown may vary in different countries, nevertheless, it is clear from Figure 1.2 that CAPLUS covers a much broader range of disciplines than that is of interest just to chemistry schools, and that in combination, these two databases cover nearly all disciplines in the sciences very extensively!

Another aspect of coverage is that CAPLUS currently contains data for more than three million records from patents. For the publication year 2000, almost one-third of the records in the database are from patents and more than 40 % of the new organic substances registered in the CAS substance database are from patent documents. No search is complete without paying attention to the very extensive patent literature!

In addition to breadth of coverage, SciFinder Scholar has many other features including currency (the major CAS databases are updated daily and MEDLINE is updated four times a week), links to full text patents and to electronic journals, options to links to in-house collections, unique search options including structure and reaction searches, and a search interface that guides the searcher through possibilities. With respect to the secondary literature, SciFinder Scholar currently provides solutions to most of the information needs of the staff and students in the sciences in a unique way, and certainly in a way not matched by any other single desktop information tool.

1.4 Types of Databases in SciFinder Scholar

SciFinder Scholar is an interface that requires minimum training, and it certainly obviates the need for detailed knowledge of search mechanics and search commands. However, some knowledge of the databases is valuable. First, it is helpful to understand that two of the databases (CAPLUS and MEDLINE) are bibliographic databases, three (REGISTRY, CHEMLIST, and CHEMCATS) are substance databases, and the last (CASREACT) is a chemical reaction database.

Some entries in bibliographic databases are text from authors and some are text from indexers. Consequently, *it helps if the searcher first thinks how an author may have described the research in the title and abstract*. In practice, there is little consistency among authors, and Chapter 2 discusses how SciFinder Scholar manages this issue primarily through automatic application of synonyms and truncation and searching for terms either within the same sentence or within the whole record.

It also helps if the searcher considers *how an indexer may have applied the index policies* of the database producer to enter text into the record. In this chapter, basic indexing principles are described, although scientists should continue to learn about indexing by careful examination of actual records. Chapter 2 discusses how SciFinder Scholar offers searchers a number of unique options to take advantage of indexing through the analysis of records by index term, supplementary term, section codes, CAS Registry Numbers, CAS Roles, and MEDLINE headings, and allowable qualifiers.

1.5 CAS Bibliographic Database (CAPLUS)

Up-to-date information on the content of CAPLUS is available at *http://www.cas.org/casdb.html*. In broad terms, the database contains more than 19 million records for original publications from a variety of primary sources. While the major source is journal articles (74 %), there is an extensive coverage of patents (16 %) and conferences (5 %); among the remaining document types are dissertations, reports, and books.

Records from about 8000 journals are included. Generally, only records from those original articles that contain information related to ‘chemistry’ are fully indexed, but 'chemistry' is *very broadly defined* by CAS, and the outcome is partly illustrated in Figure 1.2! Further information on the coverage is available through the 'more than chemistry' brochures, which may be obtained at *http://www.cas.org/SCIFINDER/SCHOLAR*.

Since 1995, CAS has included entries for all articles in approximately 1350 core journals (*http://www.cas.org/sent.html*), and details of the patents

Bibliographic Information

Spectroscopy and photosensitization of sapphyrins in solutions and biological membranes. Roitman, Leonid; Ehrenberg, Benjamin; Nitzan, Yeshayahu; Kral, Vladimir; Sessler, Jonathan L.. Dep. Phys., Bar Ilan Univ., Ramat-Gan, Israel. Photochem. Photobiol. (1994), 60(5), 421-6. CODEN: PHCBAP ISSN: 0031-8655. Journal written in English. CAN 122:26806 AN 1995:202539 CAPLUS

Abstract
A spectroscopic and photophys. study of three new sapphyrin mols. is presented. The sapphyrin backbone that was derivatized to make them water sol. possesses an absorption band around 700 nm, a desired property for biol. photosensitization. We studied the absorption and fluorescence spectra, from which evidence for aggregation in solvents of different polarities was obtained. The extent of aggregation is correlated with the nature of the attached moiety. The abs. quantum yields of singlet oxygen prodn. were measured, with 1,3-diphenylisobenzofuran as a model target, and were 0.13-0.18 in ethanol. The binding consts. to liposomes and to cells were detd. spectroscopically and were found to correspond to the hydrophobicities of the compds., with an addnl. effect, ascribed to the sugar moiety, which was found in the case of one of the sapphyrins. The efficiency of photodamage to Staphylococcus aureus by sapphyrins and hematoporphyrin was equiv., on the basis of cells killed per μg of sensitizer in the incubation mixt.

Indexing -- Section 8-3 (Radiation Biochemistry)
Cell membrane
Laser radiation
Membrane, biological
Photodynamic action
Photosensitizers
Staphylococcus aureus
(sapphyrins spectroscopy and photosensitization in solns. and biol. membranes)

Porphyrins
Role: BAC (Biological activity or effector, except adverse); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(sapphyrins, sapphyrins spectroscopy and photosensitization in solns. and biol. membranes)

158012-18-1
Role: BAC (Biological activity or effector, except adverse); FMU (Formation, unclassified); BIOL (Biological study); FORM (Formation, nonpreparative)
(sapphyrins spectroscopy and photosensitization in solns. and biol. membranes)

159858-86-3
159858-87-4
Role: BAC (Biological activity or effector, except adverse); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(sapphyrins spectroscopy and photosensitization in solns. and biol. membranes)

7782-44-7, Oxygen, biological studies
Role: BAC (Biological activity or effector, except adverse); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(singlet; sapphyrins spectroscopy and photosensitization in solns. and biol. membranes)

Supplementary Terms
sapphyrin photosensitization biol membrane

Figure 1.3 Sample record from CAPLUS. Copyright the American Chemical Society and reproduced with permission.

covered are available at *http://www.cas.org/EO/caspat.html*. While no single database contains all the world's scientific journal and patent literature, the users of SciFinder Scholar may be very confident that in CAPLUS, they are searching one of the world's most comprehensive sources of scientific information!

An example of the content of a single record is shown in Figure 1.3. After the title and bibliographic information, the records contain the abstract, followed by the indexing. Titles and abstracts are usually those written by authors, although in some cases (most notably for patents) the titles and abstracts may be rewritten by indexers to better reflect the technical aspects of the original article.

One of the greatest benefits of searching databases is that users may take advantage of the indexing added by the database producer. The advantages of indexing are many, including *inter alia*:

- *systematic terminology* so that the different terms used by authors to cover a particular topic may effectively be searched under a consistent term for that topic. In this way, use of a few index terms will produce more comprehensive answer sets;
- *precise entries into topics*, since often the database producer will add the index term only if it relates to a focus of the original research; and
- *additional entries to the record*, for example, not only does the systematic indexing of chemical substances enable precision searches to be accomplished but also important substances mentioned in the article will be entered in the indexing even though no mention of them may be made in the title or abstract written by the author.

Most database producers also arrange their indexing in a hierarchy, so that searchers may easily go from one index term to terms that more broadly, or more precisely, describe the topic. The CAS hierarchy of terms is included in the CA Lexicon®, and part of the lexicon for the *Staphylococcus* genus is shown in Figure 1.4. At this stage, it helps if SciFinder Scholar users note that in some on-line systems, searches for the *Staphylococcus* genus may be conducted more generally through successive levels from Bacillaceae, to gram-positive bacteria, to Bacteria. Alternatively, the searcher may search more specifically through individual species such as *Staphylococcus aureus* and *Staphylococcus capitis*. However, searching with the aid of a thesaurus is quite specialized and although at present SciFinder Scholar does not automatically search hierarchies, it does have a number of ways to take advantage of index terms.

```
=> E STAPHYLOCOCCUS/CT+ALL
E1       11126   BT4  Bacteria (Eubacteria)/CT
E2        1043     BT3  Gram-positive bacteria (Firmicutes)/CT
E3           0       BT2  Low G+C gram-positive bacteria/CT
E4          66         BT1  Bacillaceae/CT
E5        6115           -->  Staphylococcus/CT
                 HN  Valid heading during volume 66 (1967)to present.
E6                        UF   Staphylococci/CT
E7                        UF   Staphylococcosis/CT
E8          77            NT1  Coagulase-negative Staphylococcus/CT
E9           7            NT1  Peptococcus saccharolyticus/CT
E10          0            NT1  Staphylococcus afermentans/CT
E11          4            NT1  Staphylococcus arlettae/CT
E12          0            NT1  Staphylococcus auranticus/CT
E13      16322            NT1  Staphylococcus aureus/CT
E14          1              NT2  Staphylococcus aureus altilis/CT
E15          7              NT2  Staphylococcus aureus anaerobius/CT
E16          6              NT2  Staphylococcus aureus aureus/CT
E17          1              NT2  Staphylococcus aureus gallinae/CT
E18          3              NT2  Staphylococcus aureus hemolyticus/CT
E19         17            NT1  Staphylococcus auricularis/CT
E20         75            NT1  Staphylococcus capitis/CT
E21          3              NT2  Staphylococcus capitis capitis/CT
E22          3              NT2  Staphylococcus capitis ureolyticus/CT
E23         18            NT1  Staphylococcus caprae/CT
E24        215            NT1  Staphylococcus carnosus/CT
E25          1              NT2  Staphylococcus carnosus utilis/CT
```

Figure 1.4 Example of the CA Lexicon in the CAPLUS file on STN for the index heading *Staphylococcus*. The lexicon gives broader terms (BT), and narrower terms (NT) in the hierarchy, and the number of records in the database for the specific index terms. (Note: only the first 25 of the 80 hierarchical terms are shown here.) Copyright the American Chemical Society and reproduced with permission.

CAS indexing appears in three main parts. First the database is divided into 80 sections (*http://www.cas.org/PRINTED/sects.html*) and each record is given the section and subsection number that is considered to be most relevant to the overall content of the article. For the record in Figure 1.3, the Section Code is 8-3 (Radiation Biochemistry). These broad section codes may be useful for the refinement of answer sets, and the analysis and refinement by section codes are discussed in Chapter 2, Section 2.5.4.

The main CAS indexing, which immediately follows the listing of the section codes, consists of index headings and CAS Registry Numbers, and each of these has text-modifying phrases (which are included in parentheses).

Although a full list of indexing terms is available and the information processionals need to understand all the levels of indexing to achieve optimal searches through on-line networks, it helps if SciFinder Scholar users understand only some basic principles.

1.5.1 Index Terms: Subjects and Animal, Plant, and Microorganism Species

More than 25 000 subject index terms for subjects are used, and a guide is available at *http://www.cas.org/terms/vocab.html*. Subject terms are entered in records in which a major aspect of the original article relates to the CAS index policy for this term. In Figure 1.3, the five subject index terms are cell membrane; laser radiation; membrane, biological; photodynamic action; and photosensitizers. The CAS also systematically enters names of species when the original article describes something new about the species. More than 120 000 species' names are used, and, for example, the species' index term listed in Figure 1.3 is *Staphylococcus aureus*.

Analysis of answers by index terms is easily achieved in SciFinder Scholar and offers another important option for the refinement of answers (Chapter 2, Section 2.5.4). The searcher need not know the indexing in advance, as Scholar displays the indexing in a graphical format once initial answers have been obtained. Remember that index terms are systematic entries that may be used to help obtain more precise answers or to help with identifying terms to enter into search queries; so, part of the 'scientific experiment' involved with searching for information requires careful scrutiny of indexing.

1.5.2 Index Terms: Substance Class Headings

The CAS has defined many substance class headings and these are entered when the original article describes something new about the substance class or about a number of specific substances that are described by this class. For example, if reactions of a variety of individual alkenes are mentioned in the original text, the substance class heading 'Alkenes' will be used. (In Figure 1.3, the substance class heading 'Porphyrins' is used.)

Substance class headings always have *additional* levels of indexing through CAS Roles (see Section 1.5.4), and in Figure 1.3, the CAS Roles, namely, BAC, THU, BIOL, and USES are added to the heading 'Porphyrins' to indicate more precisely the aspect of research on porphyrins that is included in the original article.

1.5.3 Index Terms: Substances

The CAS Registry Numbers underpin the CAS Registry System, and the only complete source of CAS Registry Numbers is the REGISTRY database (Section 1.7), which may be searched in SciFinder Scholar in a number of ways (Chapter 3). Each unique substance is given a unique CAS Registry

Number and when something new is reported on the substance in the original article, the CAS Registry Number is added as an index term in the database record. Not every substance in the original article is indexed and indeed some key CAS indexing policies relating to substances are indicated in Table 1.3.

Table 1.3 Some indexing policies for substances in CAPLUS

Indexing policy	Example	Implications
The CAS Registry Number is indexed only if something new is reported for the substance in the original article	Particularly, in the introduction section, many original articles may summarize previously known information. However, unless something *new* is reported for the substance, the CAS Registry Number will not be indexed	At times, it may be advisable to search for some common names of substances as well as for the CAS Registry Numbers
The substance is indexed as precisely as possible	For example, the CAS Registry Number for morphine will be listed if the original article refers to morphine, but the CAS Registry Number for morphine sulfate will be listed if the article refers to the sulfate. Similarly, the CAS Registry Number for potassium will be listed if the article refers to potassium (e.g. potassium levels in blood), but for potassium ion if the article refers to K^+	In some cases, a few CAS Registry Numbers (e.g. the parent base and all its salts) may be needed to cover a 'substance'
Since 1985, if a simple name for the substance is given by the author, it may be included after the CAS Registry Number, although no attempts are made to apply systematic nomenclature	For example, 'oxygen' is added after the CAS Registry Number 7782-44-7 in Figure 1.3, but the other CAS Registry Numbers refer to substances with complex names and no name term is added	CAS Registry Numbers should nearly always be used in search terms for substances
If the original report refers to a number of substances of a certain class, then the substance class heading also is indexed	For example, the substance class heading 'Porphyrins' is used in Figure 1.3	General information on substance classes may be searched effectively through the substance class index headings

CAS Registry Numbers are precise and comprehensive search terms for substances and should nearly always be included in the search for substances. Search results should always be checked to see that the CAS Registry Number has been used.

The four CAS Registry Numbers in Figure 1.3 are 158012-18-1, 159858-86-3, 159858-87-4, and 7782-44-7, but only the last one has a nomenclature term. It is thus apparent that only searches using the first three registry numbers will retrieve the substances indexed in this record, which clearly has important information relating to these substances!

1.5.4 Index Terms: CAS Roles

All CAS Registry Numbers and substance class headings are followed by CAS Roles (for the list of roles, see Appendix 2). Roles have been assigned by the indexers since October 1994, while roles in the records in CAPLUS before that time have been assigned 'algorithmically'. That is, the roles have been assigned by computer-based searches that involved combinations of searches in section codes, controlled terms, and keywords. Note that searching with CAS Roles is a precision tool and is very useful for focusing on specific studies, particularly when large numbers of records occur for individual substances. The use of CAS Roles in searching through SciFinder Scholar is described in Chapter 5, Section 5.2.

1.5.5 Index Entries: Text-Modifying Phrases

Text-modifying phrases are terms in parentheses that follow the index headings or CAS Registry Numbers. They are often terms from the original article that relate most directly to the index heading, so may be considered as author-related terms that qualify the index heading. The inclusion of these terms is significant since one of the features of SciFinder Scholar is that, *inter alia*, the user may choose answers in which the concepts searched are 'closely associated'. In general, SciFinder Scholar defines terms to be 'closely associated' when they appear in the title, in a single sentence in the abstract, or within a single index term and its text-modifying phrase. The assumption is that terms that are 'closely associated' (rather than anywhere in the reference) are more directly related, and so the inclusion of text-modifying phrases within the index heading provides an important level of precision in the choice of answer sets.

In cases in which text-modifying phrases are common to a number of index headings, the index headings are grouped and the single text-modifying phrase is applied. So, in Figure 1.3, the text-modifying phrase '(sapphyrins spectroscopy and photosensitization in solns. and biol. membranes)' relates to (i.e. is 'closely associated' with) *each* of the six index headings before it.

1.5.6 Index Entries: Subheadings

The first part of the text-modifying phrase may contain a subheading. For example, in Figure 1.3, the entries 'Porphyrins, sapphyrins' and 'Oxygen, singlet' appear. At present, it is sufficient simply to note that these are subheadings, although as further functionality is introduced into SciFinder Scholar, these subheadings may become significant as they form part of the index hierarchy.

Sometimes, there may be a number of subheadings and users initially may not completely understand the entry. Since subheadings are entered first in the text-modifying phrase, it may be necessary to read the sentence in a different way. So, the entry:

IT 7782-44-7, Oxygen, biological studies
RL: BIOL (Biological study)
(singlet, formation of, porphyrins photosensitization of, after laser irradn. in aq. and alc. solns.)

is to be 'read' as 'formation of singlet oxygen by porphyrin photosensitization after laser irradiation in aqueous and alcoholic solutions'.

1.5.7 Index Entries: Supplementary Terms

Supplementary terms are natural language words or phrases that are entered by the indexers to provide key additional information about the content of the document. Generally, they relate to the authors' terminology, and although they are not controlled vocabulary terms, they may be particularly useful for refining answer sets (see Chapter 2, Section 2.5.4).

In a manner somewhat similar to that in which text-modifying phrases are often author-related terms that expand on the index term, supplementary terms may be considered as index-related terms that expand on the terms used by the authors in titles. So, supplementary terms and terms in titles are relatively precise terms.

A supplementary term in Figure 1.3 is 'biol' and, indeed, it is noted that throughout Figure 1.3, there are a number of abbreviated terms. It is very important to search for abbreviations, but as discussed in Chapter 2, Section 2.3.2, SciFinder Scholar automatically searches for the abbreviations when the corresponding full term is entered in the search!

Indexing is subject to a number of policies, and users may learn basic policies simply by looking at and interpreting records. What is important to realize is that the records are written in part by authors and in part by indexers, and the search strategies should take advantage of both the parts. The implications of this are discussed in Chapter 2.

1.6 NLM Bibliographic Database (MEDLINE®)

The MEDLINE database contains more than 12 million references to journal articles in life sciences with a concentration on biomedicine. It covers 4300 journals published since 1958 (*http://www.nlm.nih.gov/pubs/factsheets/jsel.html*) and is updated four times a week. A Fact Sheet, which summarizes the content of the database, is available at *http://www.nlm.nih.gov/pubs/factsheets/medline.html*.

An example of the content of a single record is shown in Figure 1.5. After the title and the bibliographic information, most records contain the abstract as presented in the original article. Next follows the indexing. A key to indexing in MEDLINE is the Medical Subject Headings (MeSH®), which is outlined in another Fact Sheet – *http://www.nlm.nih.gov/pubs/factsheets/mesh.html*. There are more than 19 000 main headings that are constructed into a thesaurus that links broader, narrower, and related terms in the hierarchy.

The structure of the thesaurus is detailed elsewhere – *http://www.nlm.nih.gov/mesh* – and the full thesaurus may be downloaded from *http://www.nlm.nih.gov/mesh/filelist.html*. However, a quick appreciation of what is involved may be seen through Figure 1.6, which shows the thesaurus for the main heading 'Staphylococcus' in MEDLINE on the STN network. In particular, note the BT at different hierarchical levels (Broader Term 1 Micrococcaceae, Broader Term 2 gram-positive Cocci etc.), the NT (Staphylococcus aureus and Staphylococcus epidermidis), and the definition of the main heading.

SciFinder Scholar currently does not allow searches based on MEDLINE thesaurus capabilities, but scientists may download the thesaurus and use it as a guide for searching broader or narrower terms. Certainly, MeSH® terms provide important entries into related documents, and experienced users of SciFinder Scholar will get enhanced recall if they are familiar with the terms related to their research area.

In a manner similar to the way CAS Roles qualify the CAS Registry Numbers and substance index headings in CAPLUS, most index terms in

Bibliographic Information

Spectroscopy and photosensitization of sapphyrins in solutions and biological membranes. Roitman L; Ehrenberg B; Nitzan Y; Kral V; Sessler J L Department of Physics, Bar Ilan University, Ramat-Gan, Israel. PHOTOCHEMISTRY AND PHOTOBIOLOGY (1994 Nov), 60(5), 421-6. Journal code: P69. ISSN:0031-8655. United States Journal; Article; (JOURNAL ARTICLE) written in English. AN 95098899 MEDLINE

Abstract
A spectroscopic and photophysical study of three new sapphyrin molecules is presented. The sapphyrin backbone that was derivatized to make them water soluble possesses an absorption band around 700 nm, a desired property for biological photosensitization. We studied the absorption and fluorescence spectra, from which evidence for aggregation in solvents of different polarities was obtained. The extent of aggregation is correlated with the nature of the attached moiety. The absolute quantum yields of singlet oxygen production were measured, with 1,3-diphenyl isobenzofuran as a model target, and were 0.13-0.18 in ethanol. The binding constants to liposomes and to cells were determined spectroscopically and were found to correspond to the hydrophobicities of the compounds, with an additional effect, ascribed to the sugar moiety, which was found in the case of one of the sapphyrins. The efficiency of photodamage to Staphylococcus aureus by sapphyrins and hematoporphyrin was equivalent, on the basis of cells killed per microgram of sensitizer in the incubation mixture.

Controlled Terms
Check Tags: Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.

Escherichia coli: ME, metabolism
*Liposomes
Oxygen
*Photosensitizing Agents: CH, chemistry
Photosensitizing Agents: ME, metabolism
*Porphyrins: CH, chemistry
Porphyrins: ME, metabolism
Solutions
Spectrometry, Fluorescence
Spectrophotometry, Ultraviolet
Staphylococcus aureus: ME, metabolism

Registry Numbers
17778-80-2 (singlet oxygen)
7782-44-7 (Oxygen)

Chemical Names
0 (Liposomes)
0 (Photosensitizing Agents)
0 (Porphyrins)
0 (Solutions)

Figure 1.5 Sample record from MEDLINE.

MEDLINE are qualified, and the qualifiers present in records for the heading 'Staphylococcus' are given in the AQ field in Figure 1.6. A full list of allowable qualifiers is available at *http://www.nlm.nih.gov/mesh/topcat.html*, although to assist users the qualifiers and the acronym are always present in the actual database record. For example, in Figure 1.5, the heading 'Photosensitizing Agents' is qualified both by CH (chemistry) and ME (metabolism).

In addition to the index headings and allowable qualifiers, the indexers add an asterisk (*) to those index headings considered to be key terms related to the article. Records in MEDLINE also may contain CAS Registry

```
=> E STAPHYLOCOCCUS/CT+ALL
E1            0   BT5   B Organisms/CT
E2        50112    BT4   Bacteria/CT
E3         3568     BT3   Gram-Positive Bacteria/CT
E4          430      BT2   Gram-Positive Cocci/CT
E5          201       BT1   Micrococcaceae/CT
E6        16866        -->   Staphylococcus/CT
E7        37865        MN    B3.510.400.500.846./CT
                        DC    an INDEX MEDICUS major descriptor

NOTE A genus of gram-positive,facultatively anaerobic,coccoid
bacteria. Its organisms occur singly, in pairs, and in tetrads and
characteristically divide in more than one plane to form irregular
clusters. Natural populations of Staphylococcus are membranes of
warm-blooded animals. Some species are opportunistic pathogens of
humans and animals.
INDX  infection = STAPH INFECTIONS on data form;
     staphylococcal clumping factor = COAGULASE (see X refs there);
     DF: STAPH
AQ    CH CL CY DE EN GD GE IM IP ME PH PY RE UL VI
MHTH  NLM 1966

E8            0         UF    STAPH/CT
E9        19955         NT1   Staphylococcus aureus/CT
E10        2714         NT1   Staphylococcus epidermidis/CT
```

Figure 1.6 Example of the thesaurus in the MEDLINE file on STN. The thesaurus gives broader terms (BT), narrower terms (NT), a definition for the term (NOTE), the allowed qualifiers (AQ), and used for terms (UF), and thus suggests many search opportunities.

Numbers, chemical names, and chemical terms. Note that the CAS Registry Numbers in Figure 1.5 do not correspond exactly with those used for the same original article in CAPLUS (Figure 1.3), but this simply reflects the different indexing policies of the two organizations. Although CAPLUS contains more than 33 million CAS Registry Numbers, only slightly more than 56 000 of these have listings in MEDLINE. The implications of this are discussed in Chapter 5, Section 5.2.

This section is intended to give searchers just an outline of MEDLINE indexing and to point to additional resources that may help those who wish to use SciFinder Scholar at a very advanced level. In summary, it is generally sufficient for users to understand that an index entry:

*Photosensitizing Agents: CH, chemistry

indicates that the MeSH® heading is ‘Photosensitizing Agents’, the * indicates a key heading for the record, and ‘CH, chemistry’ is a qualifier that more specifically links the nature of the research to the MeSH® heading.

1.7 CAS Substance Database (REGISTRY)

While bibliographic databases have text from authors and indexers, chemical substance and chemical reaction databases are produced entirely by indexers. Consequently, it helps if searchers understand the index policies that have been applied, and some of these are discussed below (and in more detail in Chapters 3 and 6).

Up-to-date information on REGISTRY is available from *http://www.cas.org/casdb.html*. The database currently contains more than 33 million substance records, which include 18.7 million organic and inorganic substances and 14.4 million sequences. An example of a record in SciFinder Scholar is shown in Figure 1.7.

After the CAS Registry Number, the current CAS systematic name is given, followed by former CAS names, common usage names, and trade names. These complete names may be searched through SciFinder Scholar (see Chapter 3, Section 3.6 and Chapter 5, Section 5.2). The print version of Chemical Abstracts is indexed in five-year collective indexes and the terms 9CI and 8CI in the name fields in Figure 1.7 refer to the ninth and eighth Collective Index, respectively. They are of no significance with respect to search strategies.

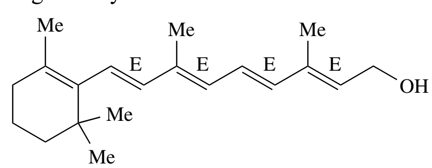
Most substance records in REGISTRY list files on the STN network that contain the CAS Registry Number for the substance, and at times, these specific databases may yield additional information. The files automatically included in SciFinder Scholar are CAPLUS (the CAS bibliographic database), CASREACT (the CAS chemical reaction database), CHEMCATS (the CAS commercial substances database), CHEMLIST (the CAS regulated chemicals database), and MEDLINE (the NLM bibliographic database). When the substance is commercially available or is included in the regulated chemicals list, the SciFinder Scholar record for the substance contains a link to the record in the appropriate database.

The indexing of substances is subject to a number of issues, which may apply either to the class of substance (e.g. salts, polymers, alloys, mixtures) or to the nature of the bonding in the molecule (e.g. issues such as resonance, σ- and π-bonding need to be addressed). SciFinder Scholar automatically handles most of these issues, although there are many complications and the user is well advised to study the basic aspects of the indexing of substances. With the aid of examples, Appendix 5 explains the most commonly encountered issues.

Registry Number:	68-26-8
CA Index Name:	Retinol (9CI)
Other Names:	Retinol, all-trans- (8CI); (all-E)-3,7-Dimethyl-9-(2,6,6-trimethyl-1-cyclohexen-1-yl)-2,4,6,8nonatetraen-1-ol; β-Retinol; 2,4,6,8-Nonatetraen-1-ol, 3,7-dimethyl-9-(2,6,6-trimethyl-1-cyclohexen-1-yl)-, (all-E)-; A-Mulsal; A-Vi-Pel; Acon; Afaxin; Agiolan; Agoncal; Alcovit A; all-trans-Retinol; all-trans-Retinyl alcohol; all-trans-Vitamin A; all-trans-Vitamin A alcohol; all-trans-Vitamin A1; Alphalin; Alphasterol;. . ... (*additional names follow*)
Formula:	C20 H30 O
STN Files:	CAPLUS, AGRICOLA, ANABSTR, BEILSTEIN, BIOBUSINESS, BIOSIS, BIOTEHNO, CA, CABA, CAOLD, CASREACT, CBNB, CEN, CHEMCATS, CHEMINFORMRX, CHEMLIST, CIN, CSCHEM, CSNB, DDFU, DETHERM, DRUGU, EMBASE, GMELIN, HSDB, IFICDB, IFIPAT, IFIUDB, IPA, MRCK, MSDS-OHS, NAPRALERT, NIOSHTIC, PHAR, PIRA, PROMT, RTECS, SPECINFO, TOXLINE, TOXLIT, ULIDAT, USAN, USPATFULL, VETU

(Additional Information is available through STN International. Contact your information specialist, a local CAS representative, or the CAS Help Desk for Assistance)

Deleted Registry Number(s):	13123-33-6, 17104-91-5, 5979-23-7
	Double bond geometry as shown.



Commercial Sources
Regulated Chemicals Listing
∼5732 References

Figure 1.7 Sample record from REGISTRY. (Note: only some of the alternative names are given in this display.) Copyright the American Chemical Society and reproduced with permission.

1.8 CAS Regulatory Information Database (CHEMLIST®)

In order for national authorities to have a mechanism to regulate trade in chemicals, many countries require companies to register the substances

Accession Number: 323 CHEMLIST
CAS Registry Number: 68-26-8
Chemical Name
Retinol (English, French, German, Spanish) (TSCA, DSL, EINECS, AICS)
Vitamin A (ENCS)
Vitamin A alcohol (ENCS)
(all-E)-3,7-Dimethyl-9-(2,6,6-trimethyl-1-cyclohexene-1-yl)-2,4,6,8-nonatetraen-1-ol (ECL)
(all-E)-3,7-Dimethyl-9-(2,6,6-trimethyl-1-cyclohexen-1-yl)-2,4,6,8-nonatetraen-1-ol
β-Retinol
2,4,6,8-Nonatetraen-1-ol, 3,7-dimethyl-9-(2,6,6-trimethyl-1-cyclohexen-1-yl)-, (all-E)-
A-Mulsal ….

File Segment
AUSTRALIA: AICS; CANADA: DSL; EEC: EINECS; JAPAN: ENCS; KOREA: ECL; USA: FDA, STATE, TSCA

Confidentiality Status
Public
Regulatory List Number
EINECS No.: 200-683-7
ENCS No.: 9-1029; 9-1053
ECL Serial No.: KE-11884

Inventory Status

On TSCA Inventory	January 2000 Inventory Tape.
On DSL	Supplement to Canada Gazette, Part I, January 26, 1991.
On EINECS	Annex to Official Journal of the European Communities, 15 June 1990.
On ENCS	Japanese Gazette. Contained within class: Low Molecular Carbo-monocyclic Organic Compounds.
On AICS	Australian Inventory of Chemical Substances, June 1996 Ed.

FDA Priority-Based Assessment of Food Additives
FDA Priority-Based Assessment of Food Additives
Priority-Based Assessment of Food Additives (PAFA) File, FDA Center for Food Safety and Applied Nutrition (CFSAN) (1998)
Listed Name(s): Vitamin A

==== Miscellaneous Regulations ====

Hazard, Toxicology, and Use Information
Hazard, Toxicology, and Use Information
Human Data
Mutation data
Reproductive Effect (RTECS)

==== U. S. State Regulations ====
Massachusetts Right-to-Know
Massachusetts Right-to-Know
Massachusetts Substance List for Right-to Know Law (11 Apr 94); General Law C.111F, Chapter 30A (28 Jun 84); 105 CMR 670.000, Appendix A
Listed Name(s): Vitamin A

Figure 1.8 Sample record from CHEMLIST. Copyright the American Chemical Society and reproduced with permission.

before their manufacture or distribution. Such inventories are essential, for example, for the monitoring of illegal substances and for keeping track of environmental issues relating to chemicals. The primary reference point in these national inventories is the CAS Registry Number.

The CAS has built a database of regulated chemical substances from a number of national and international chemical inventories and regulatory lists. The database contains more than 230 000 chemical substances and is updated weekly. Details of the content are available at *http://www.cas.org/CASFILES/chemlist.html* and a typical record is shown in Figure 1.8.

It is neither possible nor necessary to search directly in CHEMLIST with SciFinder Scholar. Instead, the user first needs to find the chemical substance and then click on the link labeled 'Regulated Chemicals Listing'. In general, the staff and students at universities are not involved in registering chemicals with national authorities, although safe laboratory practices apply to university laboratories and the regulated substance database gives important references to legal requirements and safety issues.

1.9 CAS Chemical Catalog Database (CHEMCATS®)

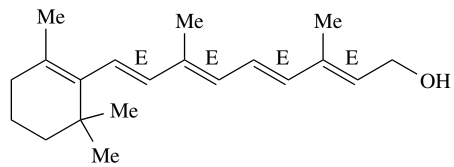
The CAS chemical catalog database *http://www.cas.org/CASFILES/chemcats.html* contains more than 2.7 million records from approximately 600 suppliers and 700 catalogs. Each record contains the catalog information for the substance (e.g. chemical and trade names, the company names and addresses) and the supplier information (e.g. pricing terms) and an example is shown in Figure 1.9.

Again, it is neither possible nor necessary to search directly in CHEMCATS. Instead, the strategy is to find the substance first. When a substance appears in one of the chemical catalogs, a link is provided (e.g. the link 'Commercial Sources' in Figure 1.7) to the chemical catalog database and it is easy for the user to scroll through the company information to identify the local supplier.

1.10 CAS Chemical Reaction Database (CASREACT®)

The reaction database contains about five million reactions selected from journal articles since 1974 – *http://www.cas.org/CASFILES/casreact.html*. A typical example in SciFinder Scholar is shown in Figure 1.10. In the reaction database, all the atoms and bonds are correlated between starting material and product, and bonds being formed or broken are tagged; however,

Catalog Name:	SIGMA
Publication Date:	25 Feb 2000
Order Number:	R7632
Chemical Name:	all trans-Retinol
Synonym:	Vitamin A
Registry Number:	68-26-8

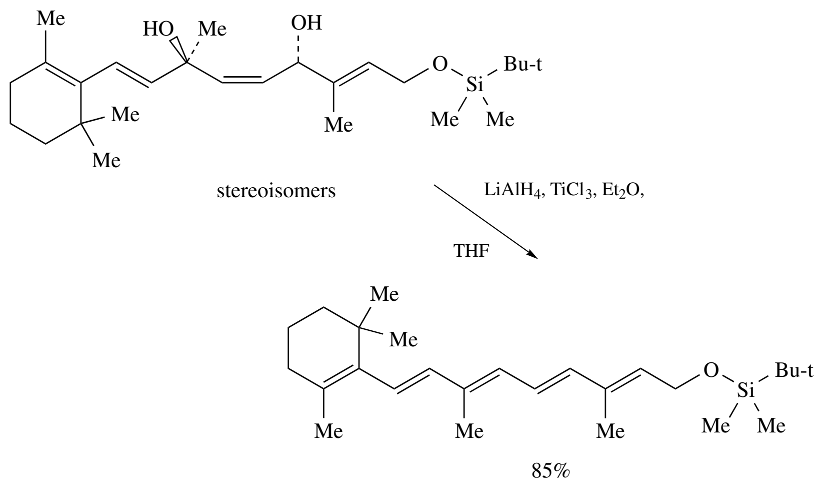


Pricing:	Quantity: 1 25 MG, Price: inquire
	Quantity: 1 100 MG, Price: inquire
	Quantity: 1 250 MG, Price: inquire
	Quantity: 1 500 MG, Price: inquire
	Quantity: 1 1 G, Price: inquire
	Notes: 1999
Company Info:	SIGMA
	P O Box 14508
	St. Louis, MO, 63178-9916 USA
	Phone: 1 800 325 3010
	314 771 5750
	Fax: 1 800 325 5052
	314 771 5757
	Email: sigma@sial.com
	Web: http://www.sigma-aldrich.com
Company Info:	SIGMA-ALDRICH de Argentina, S.A.
	Av. Pueyrredon 2446/50,Piso 5-B
	Buenos Aires, 1119 Argentina
	Phone: 54 1 807 0321
	Fax: 54 1 807 0346
	Email: rcsjr@compuserv.com

(Included also is information from other offices worldwide)

Figure 1.9 Sample record from CHEMCATS. Copyright the American Chemical Society and reproduced with permission.

these details are not seen in the reaction diagram as shown (Figure 1.10). This enables precise reactions to be retrieved easily, and this is important since the scientist may not only wish to know that a substance has been prepared but may also wish to know the preparations involving the formation of specific bonds.



Reference: J. Org. Chem., 54(11), 2620-8; 1989
6 additional hit reactions in document (click microscope to view)

Figure 1.10 Sample record from CASREACT. Copyright the American Chemical Society and reproduced with permission.

It is helpful to appreciate that only key or representative new reactions from the original documents are fully indexed and that the database includes data from 1974. Even so, it is one of the world's premier reaction databases and generally produces many more relevant answers than alternative desktop reaction databases. SciFinder Scholar offers a number of other options for searching for chemical reactions, and further details are discussed in Chapter 6.

1.11 Exercises

1.1 Spot the difference! In Figures 1.3 and 1.5, what differences are there in the:

(a) bibliographic parts of the records;

(b) abstract;

(c) subject indexing; and

(d) substance indexing.

1.2 Why is the substance 1,3-diphenylisobenzofuran not mentioned in the indexing in Figures 1.3 and 1.5, even though it is in the abstract? What are the implications?

1.3 Which part of the record in Figure 1.3 is written by the indexer?

1.4 What policies do authors apply when writing titles and abstracts, and what should authors be keeping in mind when they write these sections?

1.5 Discuss the advantages and disadvantages of searching those parts of the records written by:

(a) the author;

(b) the indexer.

1.6 Figure 1.3 contains four CAS Registry Numbers, but only one has a name associated. Why do CAS use Registry Numbers as index entries for substances and what are the implications of this for the searcher?

1.7 On the basis of the information in Figures 1.3 and 1.5, what search terms would be needed to find all records that:

(i) contained the concept photosensitization;

(ii) are written by Jonathan L. Sessler;

(iii) are listed for Bar Ilan University; and

(iv) involve the substance singlet oxygen.

(HINT: To answer this question, it is sufficient to write down the actual terms that are in the records.)

1.8 One of the search options in SciFinder Scholar is to restrict entries to those where the search terms are 'closely associated' (which usually means just within the title, within one sentence in the abstract, or within the one index term including its text-modifying phrase). While each of the following words appear somewhere in the record in Figure 1.3, which combinations of any two terms appear only 'closely associated'? What are the implications for the searcher?

sapphyrins	porphyrins	singlet oxygen
photosensitization	laser radiation	Staphylococcus aureus
hematoporphyrin	7782-44-7	

1.9 The CAS Registry Number for hematoporphyrin is 14459-29-1. The word hematoporphyrin is in the abstract in Figures 1.3 and 1.5, but the

CAS Registry Number is not indexed. Why is this so, and what are the implications for searches on the substance?

1.10 In Figure 1.7, the substance ‘retinol’ lists a number of STN files in which the CAS Registry Number appears but MEDLINE does not appear. Since retinol is an important substance involved with the chemistry of vision, is it not surprising that the CAS Registry Number 68-26-8 is not in MEDLINE? What is the ‘problem’ here? (HINT: unless you already have good knowledge of the science and of the databases, you will not be able to answer this question at this stage! However, when you logon to SciFinder Scholar, you may wish to find the CAS Registry Numbers 68-26-8 and 11103-57-4 and then think about what is happening!)