

# Chapter 1

## Bibliographic databases

Barton W. Trawick and  
Johanna R. McEntyre

National Center for Biotechnology Information, National Library of  
Medicine, National Institutes of Health, Bethesda,  
MD 20892, USA.

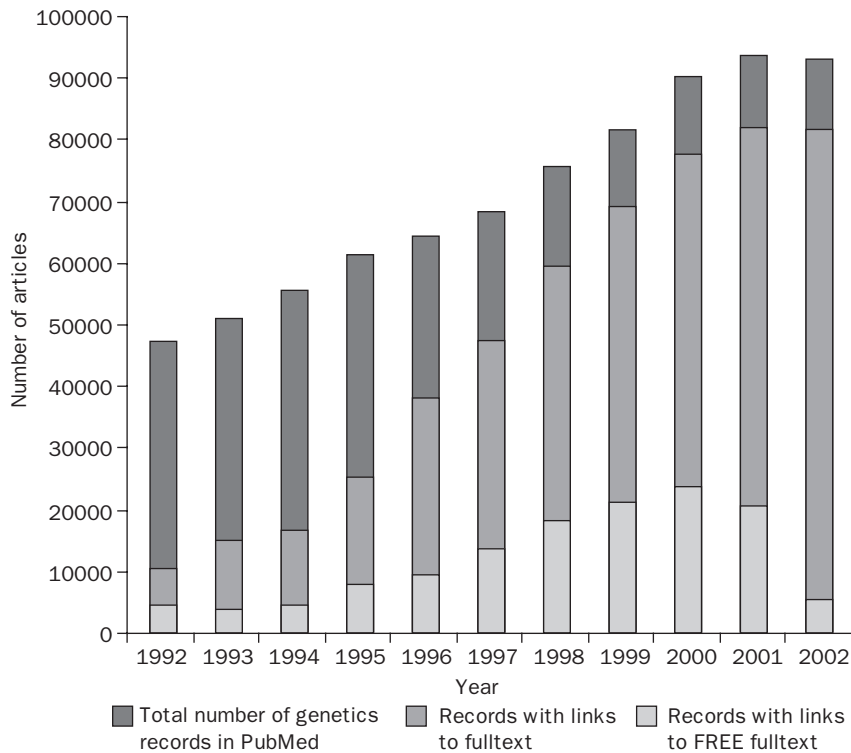
Use of the literature is fundamental to the pursuit of all knowledge. Through searching and reading, we learn what our peers are doing, develop a broader perspective on our field of interest, get ideas, and confirm our discoveries. During the course of twentieth century science, ‘the literature’ has become an expanding knowledge base that represents the collective archive of the work carried out by the international scholarly community. Recent technological advances make an increasing proportion of the literature available electronically (see *Figure 1*). This chapter offers an introductory guide for molecular biologists to stable bibliographic resources that are available over the Internet.

### 1 General introduction

The term ‘bibliographic databases’ has traditionally referred to the ‘abstracting and indexing services’ for the scholarly literature. These services focused on collecting the citation information and abstracts of research articles and making them searchable. Abstracts have been the focus for the creation of bibliographic databases because they summarize the full research article, are small enough to re-key (the only way to capture the information before electronic publishing), store, and search.

However, technological advances over the past decade have expanded the horizons of bibliographic database creation from using abstracts only to using longer pieces of text. Furthermore, the rise in use of the Internet has provided the opportunity to build online, searchable literature databases that are accessible to anyone with an Internet connection.

In response to this opportunity, publishers, libraries, and other information providers have adopted new electronic publishing technologies to develop many forms of online content. These include databases of journal abstracts, full-text



**Figure 1** The number of PubMed records that are classified under the general MESH term “Genetics” has grown from around 48,000 in 1992 to 93,000 by the end of 2002. Around 1995, a number of science, technical, and medical (STM) journals began to establish an online presence.<sup>a</sup> Since that time, the proportion of records that provide links to online full-text articles has increased. In 2001 and 2002, around 87% of “Genetics” records have links to online full-text articles; about 25% of these are freely available. (The smaller proportion of free access articles in 2002 is indicative of the common practice of publishers to delay free access for a period of time after publication.)

<sup>a</sup> Hitchcock S, *et al.* 1996. A survey of STM online journals 1990–95: the calm before the storm.  
<http://journals.ecs.soton.ac.uk/survey/survey.html>

articles, and books, as well as ‘internet-only content’ in the form of news and summaries.

For the purposes of this chapter, ‘bibliographic databases’ will be considered as ‘any large, stable, collection of primarily text-based information that is available over the Internet’. The chapter will therefore not discuss individual journal titles (although it could be argued that the online collection of articles from a single journal constitutes a small database), nor will it discuss the more popular health and medicine websites (though there are many to choose from). Further, molecular biology sequence databases frequently have

some literary or descriptive component; but if their focus is on data rather than text, it will not be discussed here. As electronic publishing is in such a state of flux, the discussion will be limited to only the most stable resources on the web.

Many of these databases require a personal subscription, a library subscription, or a site license, but several of the resources discussed are free to use.

This chapter is divided into *three* sections, based on the following types of bibliographic information described:

- 1 *Abstracts*: bibliographic databases that contain the abstracts of journal articles plus the citation information (e.g. author names and affiliations, the journal title, volume, and page numbers).
- 2 *Full-text articles*: there are now several resources available on the web that offer free access to the complete articles from life science journals.
- 3 *Books and text-rich websites*: some publishers—both traditional and new ones—are now experimenting with the online publication of textbooks, as well as new, information-rich websites.

## 2 Abstracts

When investigating a new topic area or seeking an update on a known research area, searching an online collection of abstracts of journal articles is often the first approach.

The strategy used to search abstracts databases is central to how successful you will be in finding what you are looking for (or discovering things you did not know you were looking for!). A search query that is too broad will pick up so many abstracts as to be useless, while one too specific might be too limiting for an expansive search. It takes practice to find the right balance and may require the use of Boolean search constructs and techniques such as delimiting the search by restricting it to specific fields, for example, searching only author names, or only article titles. A general introduction to the use of advanced search techniques is summarized in *Box 1*.

### **Box 1** Tips and tricks for searching bibliographic databases

#### *Boolean searching*

Boolean expressions (named in honour of the English mathematician George Boole) allow the user to combine 'AND', 'OR', and 'NOT' operators with specific search terms in order to create a more defined query. In most search engines, operators may be combined (solved in order from left to right), and parentheses '(' )' may be employed to clarify terms, group them together, and change the order in which expressions are solved.

**Box 1 (Continued)**

The following expressions:	Will return references that contain:
Watson AND Crick	<i>at least</i> both terms 'Watson & Crick'
Watson OR Crick	Either 'Watson' or 'Crick', or 'Watson & Crick'
Watson NOT Crick	'Watson' but not 'Crick'
Wilkins AND Watson AND Crick	<i>at least</i> 'Wilkins & Watson & Crick'
(Watson AND Crick) OR Wilkins	'Watson & Crick' or 'Watson, Crick & Wilkins' or 'Wilkins'
Wilkins AND (Watson OR Crick)	'Wilkins & Watson' or 'Wilkins & Crick' or 'Wilkins, Watson, & Crick'
Watson AND Crick NOT Wilkins	<i>at least</i> both terms 'Watson & Crick' with no occurrences of 'Wilkins'

*Limiting searches to fields*

In addition to using Boolean expressions to define a more specific output, it is also possible to limit individual terms to fields. For example, the term 'Crick' may be limited to 'Author Name' and 'Nature' may be limited to 'Journal Name'. The number of possible fields that are in a given database may vary, but at a minimum usually include: Author Name, Author Affiliation, Journal Name, Article Title, Publication Date, Page Number, Issue, and Volume.

*History functions*

The results of two or more searches can be combined to form a third output, or additional terms may be added to results from previous searches through the use of 'history' functions. This can be particularly useful for reducing large search results into smaller, more focused ones or for combining several different terms with a single common term. For example, independent queries for 'cancer', 'DNA repair', '1995', 'Vogelstein', 'human', and 'mouse' could be used in various permutations and combinations (linked by Boolean expressions) to form new queries. Additionally, some bibliographic databases can be customized so that useful queries can be stored for future use.

For an example of how these search techniques can be combined to search PubMed, see *Protocol 1*.

## 2.1 Databases—in all their forms

There are several abstracting and indexing services available online, many of which require a subscription. There is considerable content overlap among the major bibliographic databases, and for this reason your library is unlikely to subscribe to all of them.

When considering the use of any of these databases, it is important to make a distinction between the database itself (i.e. the physical collection of abstracts) and the access route into the information. Several of the large databases can be accessed from more than one place, because the owners of the data (i.e. the abstracts collection) lease or sell their data, or have allowed service providers to furnish a portal to the information.

*Table 1* lists the major abstracts databases for molecular biology, along with the owner of the database and a list of access points into the data. MEDLINE, for example, is one of the most widely used abstracts databases. Some of the abstracts of MEDLINE can be distributed freely (those under copyright require permission to reproduce them), so many organizations have developed clones or interfaces to MEDLINE that generate alternative portals to the same information. This can

**Table 1** Abstracts databases

Resource	Produced by	Examples of access	Free access*	URLs
PubMed/MEDLINE	The National Library of Medicine (NLM)	PubMed BioMedNet	Yes Yes	<a href="http://www.pubmed.gov">http://www.pubmed.gov</a> <a href="http://research.bmn.com/medline">http://research.bmn.com/medline</a>
ISI Citation Database (Web of Science)	Institute for Scientific Information (ISI)	Ovid BIDS	No Yes	<a href="http://www.ovid.com/">http://www.ovid.com/</a> <a href="http://www.bids.ac.uk/">http://www.bids.ac.uk/</a>
Current Contents®	Institute for Scientific Information (ISI)	Web of Science	No	<a href="http://www.isinet.com/isi/journals/">http://www.isinet.com/isi/journals/</a>
BIOSIS Previews® (comprising biological abstracts and biological abstracts/RMM®)	BIOSIS	Current Contents Connect Ovid	No No	<a href="http://www.isinet.com/isi/journals/">http://www.isinet.com/isi/journals/</a> <a href="http://www.ovid.com/">http://www.ovid.com/</a>
Pascal	Institut de l'Information Scientifique et Technique	BIOSIS Ovid	No No	<a href="http://www.biosis.org/">http://www.biosis.org/</a> <a href="http://www.ovid.com/">http://www.ovid.com/</a>
EMBASE	Elsevier Science	BIDS	Yes	<a href="http://www.bids.ac.uk/">http://www.bids.ac.uk/</a>
The Cochrane Reviews (abstracts)	The Cochrane Library	EMBASE.com Ovid The Cochrane Library	No No Yes	<a href="http://www.embase.com/">http://www.embase.com/</a> <a href="http://www.ovid.com/">http://www.ovid.com/</a> <a href="http://www.update-software.com/abstracts/crgindex.htm">http://www.update-software.com/abstracts/crgindex.htm</a>

\* In cases where access to the database is not free, consult your library for subscription information.

provide a useful addition to a publisher's website, or it may produce an interface in a language other than English.

### 2.1.1 The databases

#### 2.1.1.1 PubMed/MEDLINE

PubMed was developed at the National Center for Biotechnology Information (NCBI), within the National Library of Medicine (NLM), USA. It encompasses the over 12 million abstracts in MEDLINE, and currently covers about 4000 biomedical journals, dating back to 1966. MEDLINE abstracts have a controlled vocabulary associated with them known as Medical Subject Heading (MeSH) terms. Several terms are assigned to each MEDLINE abstract, and are used for indexing articles to provide a consistent way to retrieve information.

As well as enabling abstract searches (e.g. see *Protocol 1*), PubMed offers the following additional functions:

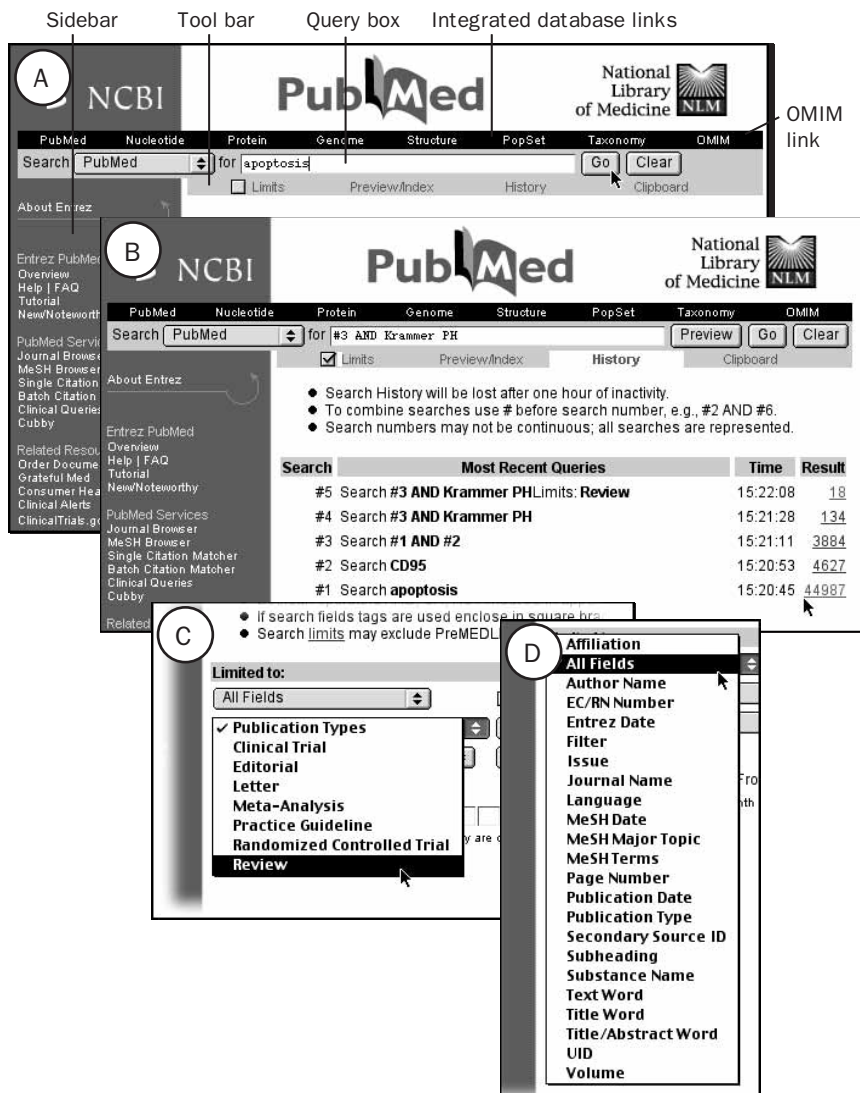
- 1 Links to biological sequence information, including data such as GenBank protein and nucleotide sequences, and macromolecular structures.
- 2 Links to the full-text of journal articles (about 4000 journals are currently linked in this way). Whether the full text can be viewed without purchasing the journal depends on the journal policy (see section below on full-text articles).
- 3 Links to 'Related articles'. For each abstract, similar articles in the database have been identified, based on a statistical analysis of words and phrases found in the abstract text. This is an easy way to expand on a PubMed search when a useful abstract has been found.
- 4 Links to resources outside of the NLM. The 'LinkOut' feature allows other providers of information, such as organism-specific databases like FlyBase, to link to related abstracts.
- 5 Links to textbooks. A new collaborative project at the NCBI is linking the content of textbooks to PubMed abstracts to serve as background information (see Section 4.2).

PubMed is primarily a biomedical database that historically has not collected abstracts from non-medical areas of molecular biology. However, more recently, the scope of PubMed has widened to include coverage of those areas, such as the plant sciences. PubMed does have the significant advantage that it can be used free-of-charge from anywhere in the world.

## Protocol 1

### Using PubMed

The PubMed page (*Figure 2(a)*) consists of the following: (1) A sidebar that contains links to PubMed information and services; (2) A query box for entering search terms; (3) A feature bar that contains links for advanced searching; and (4) Links to other integrated molecular biology databases.



**Figure 2** The PubMed search interface. (a) The PubMed page has a sidebar that links to related services and information, links to integrated databases, a query box for entering search terms, and a tool bar that contains links for advanced searching. (b) Previous searches can be viewed using the history feature. Searches can be combined through use of search numbers and Boolean operators. (c) The 'Limits' feature allows searches to be constrained to various information fields, such as Author Name or Review Article. (d) The field restrictions are found only in the 'All Fields' pull-down menu.

**Protocol 1** continued

As an example of how a PubMed search can be conducted, we will look for review articles on CD95 (or Fas, a lymphocyte receptor) and apoptosis, written by P. H. Kramer.

**Method**

- 1 Enter the search term ‘apoptosis’ in the query box and click the ‘Go’ button to the right of the box to initiate the search (*Figure 2(a)*). Conducting a search using a broad term without any field restrictions usually returns a large number of hits; in this case, more than 70,000 citations are found.
- 2 PubMed retains the most recent search term in the query box. Click the ‘Clear’ button to the right of the query box to remove the previous search term (‘apoptosis’) and replace it with the new search term ‘CD95’ and click ‘Go’. The search for ‘CD95’ returns several thousand references.
- 3 Click on the ‘History’ tab located in the feature bar. The results of the two previous searches are now displayed chronologically in a numbered list. These results may be reviewed individually by clicking on the number of returns for each query in the ‘Results’ column (*Figure 2(b)*).
- 4 Searches stored in History may be combined using Boolean operators (see *Box 1*) to form a new search. Search for references that contain both ‘apoptosis’ and ‘CD95’ by typing ‘#1 AND #2’ in the query box and clicking on the ‘Preview’ button. Selecting ‘Preview’ will display the search results in History summary format, rather than listing each article found. (Note: PubMed is case-insensitive for search terms but case-sensitive for Boolean operators: make sure ‘AND’ is in capitals.)
- 5 New queries may be combined with previous searches. To find references associated with<sup>a</sup> the name P. H. Kramer, type ‘#3 AND kramer ph’ into the query box (*Figure 2(b)*) and click ‘Go’.
- 6 Queries can be limited to various fields such as: journal name, author name, title word, MESH term, publications type, publication date (or date range), or language. To employ limits in PubMed, click on the ‘Limits’ tab in the feature bar. Limit the ‘Publication Types’ pull-down menu to ‘Review’ (*Figure 2(c)*). Check that the search term ‘kramer ph’ is still present in the query box and click ‘Go’. The result displays a summary view of all review articles associated with<sup>a</sup> P. H. Kramer that contain the query terms ‘apoptosis’ and ‘CD95’.
- 7 Limits will remain in effect for subsequent searches unless they are deselected by clicking the check box to the left of the ‘Limits’ tab in the feature bar. Additional limits are located in the ‘All Fields’ pull-down menu (*Figure 2(d)*). For a more comprehensive account on advanced searching of PubMed, consult the help documentation, listed in the sidebar on the PubMed homepage.

<sup>a</sup> When searching for authors by last name only, rather than for abstracts that merely cite the author’s last name, the search should be carried out with the field limited to



**Protocol 1** continued

Author Name. For example, a search for 'Crick' without any field limits will return abstracts that contain the word 'Crick', such as in the term 'Watson-Crick base pair'. Similarly, a search for articles in the journal 'Cell' needs to be executed with the field limited to Journal Name, otherwise the results will list any abstract that contains the word 'cell'.

### 2.1.1.2 *Web of science*

The Institute for Scientific Information (ISI) produces the 'Web of Science'—an interface to the ISI Citation Database that contains more than 5300 scientific articles, dating from 1980, that is updated weekly. The Web of Science is a subscription-based service, available from many (but not all) university libraries. The Web of Science shares some features with PubMed, such as links to biological sequence information and full-text articles, but also has some that are unique:

- 1 Links to related articles. The way in which related articles are calculated in the Web of Science differs from the related articles of PubMed. In Web of Science, the list of records related to a given article consists of papers that cite at least one source also listed in the original (parent) article, with the source that has most common citations listed first.
- 2 Links to (i) the Derwent Innovations Index, a patent database; (ii) BIOSIS Previews, a database of references to primary journal literature, meetings, and books; (iii) ISI Chemistry Server, for newly reported structural chemistry.

For many molecular biologists, one of the most valuable attributes of the Web of Science comes from the use of the citations associated with each abstract. Through the references cited within an article, it is possible to:

- (a) View the abstracts of all articles cited in the original (parent) article,
- (b) Find all articles published, since the original (parent) article, that have cited it, and
- (c) Find all the articles that have cited a particular author.

### 2.1.1.3 *Current contents*

The Web of Science also interfaces with the ISI Current Contents databases, for which a subscription is required. Current Contents used to be a paper publication, distributed weekly and consisting of the contents of recently published journals, divided into broad subject categories, such as the Life Sciences (coverage of about 1400 journals). The Current Contents database can be searched, abstracts of articles found can be viewed, and from there the table of contents of the journal issue can be displayed and browsed.

#### 2.1.1.4 *EMBASE*

EMBASE (1974–present) is a bibliographic database produced by Elsevier that covers over 4000 journals in the biomedical and pharmacological sciences. Its online presence now incorporates selected MEDLINE records, thus increasing the scope and scale of EMBASE to over 13 million abstracts. Like PubMed and Web of Science, EMBASE has links from appropriate abstracts to selected full-text articles and gene sequence information. EMBASE is available by library subscription only.

#### 2.1.1.5 *The Cochrane Abstracts*

The Cochrane Reviews is a collection of reports that collate and summarize published health care evidence on a wide range of medical disorders and conditions. The target audience is very broad, ranging from those receiving care, to those responsible for research, teaching, funding, and administration of health care at all levels.

The reports are written and maintained by international panels of clinicians, who are organized into groups on the basis of area of expertise. There are currently about 50 Collaborative Review Groups that cover areas such as breast cancer, schizophrenia, HIV/AIDS, and tobacco addiction. Once a review is written, it is checked regularly and updated as needed.

While a subscription is required to access the full Cochrane Reviews, anyone can browse or search the Cochrane Abstracts without charge. The abstracts alone are quite substantial (usually about 300–500 words). They outline the background for the study, the source data, search strategy, and criteria for inclusion, and then state the results and conclusions. The Cochrane Abstracts, while more focused on clinical trials and therapies than basic molecular biology, are a high-quality and useful adjunct for those who work on molecular biology problems with clinical applications.

#### 2.1.1.6 *BIOSIS Previews*

BIOSIS Previews is made up of two databases: Biological Abstracts, which contains about 12 million records from more than 5000 journals, and Biological Abstracts/RRM, which covers reports, reviews, and meetings—information not formally published in scientific research journals. This includes references to items from meetings, symposia, and workshops, review articles, books, book chapters, software, and US patents related to the life sciences. It covers the biological sciences, from biochemistry to zoology, and is available by subscription only.

#### 2.1.2 Access providers: BIDS and Ovid

BIDS and Ovid are companies that aggregate databases created by other organizations into convenient packages for libraries to use. BIDS may be the best-known bibliographic service for academics in the United Kingdom and Ireland. It provides access to a number of databases, some of which are freely

available, and links to full text articles via Ingenta Journals (see Section 3). Many databases and services formerly provided by BIDS, including Medline, are now provided free to UK academics via the ISI's Web of Knowledge interface (<http://wok.mimas.ac.uk>).

Your library may also use Ovid as a provider of several bibliographic databases, including BIOSIS Previews®, Current Contents®, EMBASE (Excerpta Medica Database), and MEDLINE, among others.

Database aggregators often implement databases in their own way, so the interface for searching the databases may have several features that differ from the implementations of other providers.

### 3 Full text of research articles

Most of the databases described above concern abstracts of published research articles. Although not considered traditionally as bibliographic information, no discussion of online text resources would be complete without considering the increasing availability of full-text articles.

Several thousand molecular biology journals are now available in electronic form (*Figure 1*)—most are online counterparts to paper journals, but some are online-only publications (see *Box 2* for a summary of the advantages of online articles over articles printed on paper). All can be viewed via a web browser, providing that, with a few exceptions such as the *Journal of Clinical Investigation*, you have a subscription. However, more recently, some journals have made articles from back issues freely available, and new publishing ventures that offer free access to articles are emerging (see *Table 2*).

#### **Box 2** Advantages of online journals over paper journals

##### *Searchable content*

Articles in digital format may be searched for words and phrases. Most bibliographic databases provide a search engine that allows for content matching across all entries. Once an article has been obtained, the 'Find' feature in your web browser can be used to search within the article for specific words and phrases.

##### *Hypertext links*

Online articles displayed in HTML can exploit hypertext linking to create connections between related content. Links can be made from references cited in the text, to its listing in the bibliography, or to external information sources such as PubMed abstracts, referenced citations, errata, sequence information, macromolecular structures (PDB files), or even the author's home page.

##### *Multimedia*

The content of traditional printed journals is restricted to what can be presented on paper. However, online journals are able to 'add value' to articles with movies, audio, and the inclusion of large data sets (an entire genome sequence, for instance). Additionally, use of color figures does not generally represent a higher publishing cost for online journals as it does for print journals.

**Box 2 (Continued)***Accessibility*

Electronic articles can be accessed over the Internet rather than visiting a library. This is particularly useful for those in remote locations. Downloaded electronic articles can be stored on your personal computer.

*Flexible publishing model*

Some scientific journals make their content available online before the printed copy. Some journals even provide a 'rolling model' of publication where articles are accessible online as soon as they are accepted for publication. Manuscript submission, online peer review, and access of electronic content may be provided by some journals through the Internet.

**Table 2** Online full-text journals

Resource	Produced by	No. of journals <sup>a</sup>	Free access <sup>b</sup>	URL
Science Direct	Elsevier Science	1100	No	<a href="http://www.sciencedirect.com/">http://www.sciencedirect.com/</a>
Link	Springer-Verlag	500	No	<a href="http://link.springer.de/">http://link.springer.de/</a>
Interscience	Wiley	300	No	<a href="http://www.interscience.wiley.com/">http://www.interscience.wiley.com/</a>
BioMed Central	Current Science	130	Yes	<a href="http://www.biomedcentral.com/">http://www.biomedcentral.com/</a>
Society and small publisher online journals <sup>c</sup>	Highwire Press	340	Some	<a href="http://highwire.stanford.edu/">http://highwire.stanford.edu/</a> and individual journal URLs
PubMed Central <sup>d</sup>	The National Library of Medicine	150	Yes	<a href="http://www.pubmedcentral.gov">http://www.pubmedcentral.gov</a>

<sup>a</sup> Journal figures given in round numbers. Figure represents the total number of journals in each resource; not all of these may be life science journals.

<sup>b</sup> In cases where access to the database is not free, consult your library for subscription information.

<sup>c</sup> HighWire Press enables small publishers to make their journal content available online. It is not the publisher of these journals.

<sup>d</sup> PubMed Central is an active archive for journal content; it is not a publisher.

### 3.1 Access to the full text of research articles

In the absence of a search engine that indexes a good proportion of full-text life science journals, the best route to finding full-text articles is not always obvious. As mentioned above, access to an article is only possible if you or your library has a subscription, or if the article is made freely available. Below we outline the most common and useful routes to online journal articles.

#### 3.1.1 Access through abstracts databases

Most of the databases listed in the previous section can make links between abstracts and the corresponding online full-text article. There will be a link that

leads the user seamlessly to the article if the following is true:

- 1 The journal (more specifically, the journal issue) is published online.
- 2 The publisher of the journal has agreed with the database to make the article available via this route.
- 3 You or your library subscribes to the journal, or the publisher makes the article freely available.

For example, a search of an abstracts database will result in a list of ‘hits’ consisting of the citation information for each article retrieved by the query (see *Box 1*). If the abstract satisfies points (1) and (2) above, then there will be a link to the journal publisher’s website (this may only become apparent when viewing the complete abstract rather than the citation information). Clicking on this link will take you to the full-text of the article if point (3) is satisfied. Many of the freely available articles can be found in this way, and, as an example of scope of access, about 4000 journals currently have links from PubMed abstracts to their respective articles on the publisher’s site.

### 3.1.2 Access from publisher sites

Many publishers do not collaborate with all bibliographic databases to allow access to their journals, and the most conservative may only allow access to their journals by logging-on directly to their own website. In these cases, the only way to access the full-text is through your library’s interface to the journal, or by a direct visit to the journals’ website, if you hold a personal subscription. Here we will list some of the most significant places where there is a collection of full-text articles (see also *Table 2*).

#### 3.1.2.1 *HighWire Press*

HighWire Press works with scientific societies and publishers to create online counterparts to their print journals. There are currently over 340 journals that are available at HighWire, of which about 150 now offer free access to back issues of the journal.

The period of time after which the article becomes freely available depends on the policy of the journal. Some journals, such as *British Medical Journal (BMJ)* have an immediate free-access policy (i.e. anyone can look at the most current version of *BMJ*). However, most HighWire journals operate under a delayed-release policy for free full-text articles, ranging from 2 months to 5 years, with most opting for a 1–2 year delay. In total, there are now (Spring, 2001) around 250,000 free articles available. HighWire allows a basic search across all the journals they collaborate with, although the free articles are not clearly delineated.

#### 3.1.2.2 *Individual publishers*

Many publishers have developed their own online interfaces to their journal databases. Some of the largest of these are listed in *Table 2*, although there are

many smaller collections. For all these sites there is usually some free introductory content, but the journal content is almost always available only on a subscription basis.

### 3.1.3 Archives for full-text articles

Publishing journals online is still a relatively new enterprise. Now that there is a substantial volume of information available over the Internet, the question of how to effectively archive the data and make the best use of the electronic medium for searching and linking becomes obvious.

A recent initiative called PubMed Central, based at the National Library of Medicine USA, is aimed towards creating an archive for full-text life science journal articles that can be browsed and cross-searched freely. The idea is that any journal article available via the PubMed Central site can be viewed by anyone with an Internet connection from anywhere in the world.

Currently, about 150 journals are making their content available via PubMed Central, and though small at present, the potential of this kind of initiative for the future makes PubMed Central worthy of mention.

## 4 Books and text-rich websites

While books have been less evident than journals in making the transition from paper to electronic form, a few online texts do exist, although most require a subscription or site license. A growing trend is for books to have associated websites for further information and corrections (as this book has\*). These are usually listed prominently in the book. Furthermore, as biological content on the Internet evolves, so do content-rich websites that do not fit into any traditional bibliographic mold; this category of bibliographic resource is not well-defined, so here we will discuss just two of the larger and more stable resources (see *Table 3* for URL).

### 4.1 Online Mendelian inheritance in man

Online Mendelian Inheritance in Man (OMIM) is a catalogue of human genes and genetic disorders (see *Table 3* for URL). It now contains about 15,000 records, and is authored and edited by Dr Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere. The online version has been developed by the NCBI.

The OMIM database is usually searched using the name of a genetic disorder or the name of a gene to retrieve records, and it is possible to use Boolean search constructs as well as field limitations, such as chromosome number (see *Box 1*).

**Table 3** Online books

Resource	URL
Online Mendelian Inheritance in Man	<a href="http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM">http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM</a>
Online books at NCBI	<a href="http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books">http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books</a>

OMIM does not contain any figures or graphics, but it does have links to PubMed, gene, and protein information. OMIM is now one of the databases integrated with PubMed, and can be accessed for searching by clicking the OMIM link on the PubMed search page (see *Figure 2*).

## 4.2 Online books

A project to put biomedical textbooks online, make them searchable, and integrate them with PubMed and other data resources has recently begun at the NCBI (see *Table 3* for URL). There are currently about 24 books participating in the project, which broadly cover the subject areas of basic molecular and cell biology and genetics (*Figure 3*); more books are set to become available in the near future.

The book collection may be searched directly, using a similar interface to PubMed. In addition, all PubMed abstracts have a 'Books' link; clicking on this link brings up a facsimile of the abstract with hyperlinked terms and phrases that lead to the most relevant sections of the book(s) for the linked phrase. PubMed abstracts are rich in information, but they do not explain the terms or concepts used, so linking abstracts to books as background information may help address this shortfall. The quantity and subject area of hyperlinked phrases in an abstract will depend on how much the content of the abstract overlaps with that of the books available.

While the complete contents of the book are free to use in this way, for some books it is not possible to navigate across the whole book content, from chapter to chapter. In these cases, access is limited to 'stand-alone' chapters or sections.

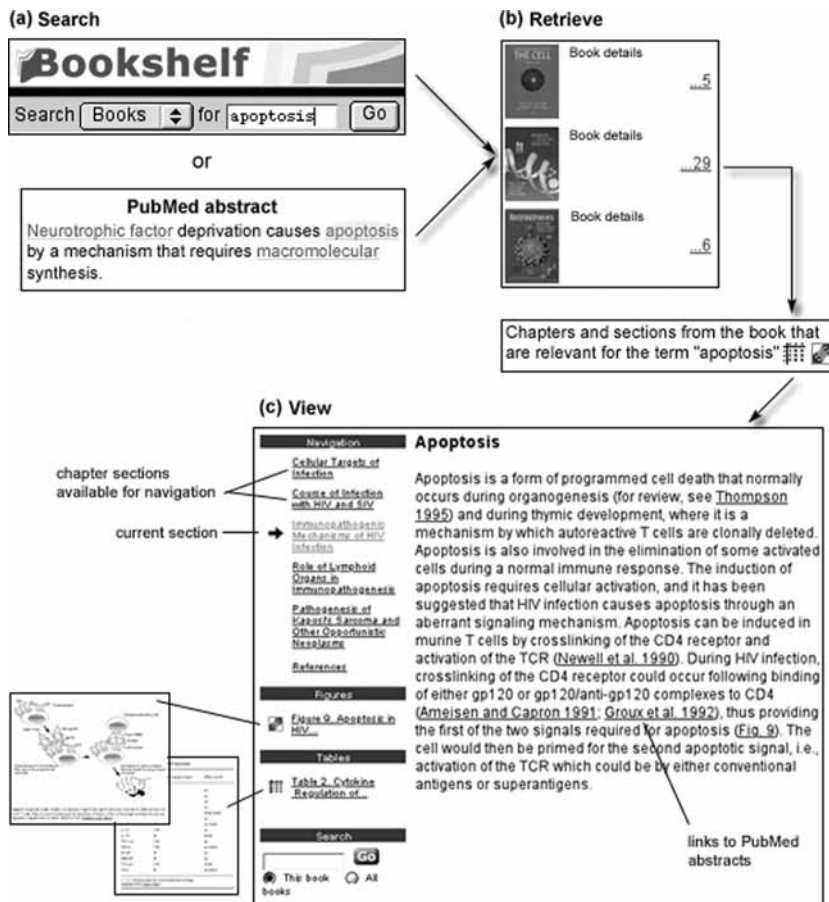
## 4.3 Text-rich websites: a word of caution

Any web search engine can also be used to search for molecular biology information. Many publishers, biotech companies, research labs, teachers, and others display information that can be browsed freely.

Information found in this way should be carefully evaluated. Be aware that anyone can publish almost anything on the Internet, so a key factor in assessing the validity of any information found is the reliability of its source. It is important to assess what qualifies the individual or organization to publish the information, and what their motivation for doing so has been. As with any literature search, the information found should be cross-checked and critically evaluated before believing.

## 5 Summary

Bibliographic information on the Internet for molecular biologists continues to grow. This chapter must really be considered a snapshot, serving as an introduction to the potential for exploring online literature resources. For this reason we have chosen to discuss only the most stable of resources, and have not discussed the specific use of any one search interface. The websites and databases discussed undergo constant evolution, and new resources are continually launched and developed. The Internet moves faster than the print world; we hope that this chapter will at least be in the same race for some time!



**Figure 3** How to access the books at NCBI. (a) All books can be searched directly from the books homepage (see Table 3 for URL), as well as indirectly, through hyperlinked phrases in PubMed abstracts. Each PubMed abstract obtained by searching PubMed has a 'Books' link. Clicking on this link displays the same abstract with some hyperlinked phrases, as shown here. (b) Executing a books search or clicking on a hyperlink within a PubMed abstract displays a summary list of books in which that term is found. The number on the right indicates the number of book sections that are relevant for the term. This link leads to a book-specific list of sections, figures, and tables. Figures and tables are indicated by the icons shown. (When less than 20 relevant book sections are found, the book summary step is omitted.) (c) The section, table or figure titles lead to the book content. The books are displayed as one chapter section per page, and it is possible to navigate around a minimum of one chapter at a time. The books contain links to the figures and tables of the book, PubMed abstracts, and in the future will be more extensively linked to molecular biology information.

### Acknowledgements

We would like to thank Kathi Canese and Edwin Sequeira for carefully reading this manuscript.