SEARCH BIBLIOGRAPHIC DATABASES: AN INTRODUCTION

CAS | STN
CAS is a division of the American Chemical Society

www.cas.org

WHAT YOU WILL LEARN

In this workbook, you will learn how to:

- Find information about the content of various databases
- Determine what steps are required for planning and conducting a search
- Use the main commands for conducting an online session

RECOMMENDED BASIC TEXT LEARNING PATH

The basic text search workbooks are intended for new users to learn the fundamentals of searching STN® or an experienced user, for a review of specific techniques.

If you are new to database searching, it is recommended that you use these workbooks in the order shown, since each workbook builds on concepts and skills covered in preceding workbooks.

1. Connect to STN
2. Work with Transcripts
3. **Search Bibliographic Databases: An Introduction**
4. Search for Multiple Terms in Bibliographic Databases
5. Evaluate and Display Search Results
6. Broaden a Search
7. Refine a Search
8. Search Author Names
9. Search Organization Names
10. Save Data
11. Access Online Help

Other STN learning paths are also available in the Learning Solutions resource center at learningsolutions.cas.org.

Workbook Convention

STN Express® is used to illustrate concepts throughout these training workbooks, although the command line techniques also apply to STN® on the Web℠.

FIND STN DATABASE INFORMATION

STN provides electronic access to more than 150 scientific and technical databases (also called files).

STN databases are created by extracting pieces of information from original, published documents such as patents, journals, and dissertations. Most STN databases contain bibliographic information, such as the document title, authors, publisher, and abstract. Some databases include information about chemical structures, reactions, and/or properties.

While most databases are in English, some databases are in other languages, such as PATDPAFULL which is in German.

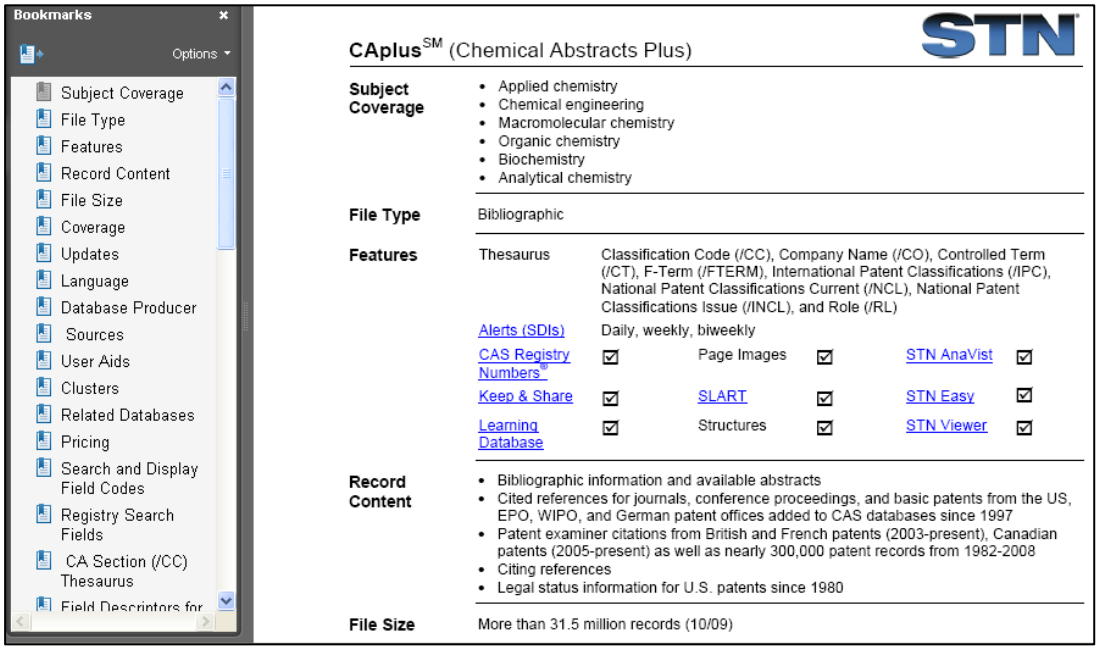
STN Database Summary Sheets (DBSS)

The Database Summary Sheets are PDF files and are used as the primary reference for learning about the content, fields, and features in a database.

Each DBSS includes information about:

- The content of the database
- Search and display fields for individual records
- Sample records

Example: Front Page of a DBSS



STN Learning Databases

Several STN databases offer learning versions in which you can practice searching prior to using the regular databases. The learning databases are indicated with the letter L preceding the name, such as LCA℠ for the CA℠ database that contains Chemical Abstracts information. Each learning database contains a static subset of records taken from the primary database.

Most STN training workbooks include practice search exercises that you can do in the LCA, CA, or CAplus℠ databases. The difference between the CA and CAplus database is that the CA database is updated weekly and the CAplus database is updated daily.

INTRODUCTION TO BIBLIOGRAPHIC RECORDS

STN databases contain records. In a bibliographic database, each record includes data for a single document such as a journal article.

BIBLIOGRAPHIC INFORMATION FOR A JOURNAL

CAplus records have three sections:

- Bibliographic information
- Abstract
- Indexing

The following example displays the content of a bibliographic record for a journal article in the CAplus database. It displays the name of each field, followed by the corresponding data extracted from the original document.

```
ACCESSION NUMBER:         2008:928341  CAPLUS  Full-text
DOCUMENT NUMBER:          149:302606
TITLE:                    Moromycins A and B, Isolation and Structure
                          Elucidation of C-Glycosylangucycline-Type Antibiotics
                          from Streptomyces sp. KY002
AUTHOR(S):                Abdelfattah, Mohamed S.; Kharel, Madan Kumar;
                          Hitron, John Andrew; Baig, Irfan; Rohr, Jurgen
CORPORATE SOURCE:         Department of Pharmaceutical Sciences, College of
                          Pharmacy, University of Kentucky, Lexington, KY,
                          40536-0082, USA
SOURCE:                   Journal of Natural Products (2008), 71(9), 1569-
                          1573
                          CODEN: JNPRDF; ISSN: 0163-3864
PUBLISHER:                American Chemical Society-American Society of
                          Pharmacognosy
DOCUMENT TYPE:            Journal
LANGUAGE:                 English
```

Based on your settings, STN can display abbreviations, or field codes, for each field.

The same record is shown here using the field codes.

```
AN   2008:928341  CAPLUS  Full-text
DN   149:302606
ED   Entered STN:  05 Aug 2008
TI   Moromycins A and B, Isolation and Structure Elucidation of
     C-Glycosylangucycline-Type Antibiotics from Streptomyces sp. KY002
AU   Abdelfattah, Mohamed S.; Kharel, Madan Kumar; Hitron, John Andrew; Baig,
     Irfan; Rohr, Jurgen
CS   Department of Pharmaceutical Sciences, College of Pharmacy, University of
     Kentucky, Lexington, KY, 40536-0082, USA
SO   Journal of Natural Products (2008), 71(9), 1569-1573
     CODEN: JNPRDF; ISSN: 0163-3864
PB   American Chemical Society-American Society of Pharmacognosy
DT   Journal
LA   English
```

Abstract

In the CAplus database, abstracts are entered in English, regardless of the original language of publication, as shown in the example below.

```
AB    Two new anticancer antibiotics of the angucycline class, moromycins A and
B, along with the known microbial metabolites saquayamycin B and fridamycin D
were isolated from the Et acetate extract of a culture broth of the terrestrial
Streptomyces sp. KY002.  The structures consist of a tetrangomycin core and
various C- and O-glycosidically linked deoxysugars. The chemical structures of
the new secondary metabolites were elucidated by 1D and 2D NMR and by mass
spectrometry.  Moromycin B showed significant cytotoxicity against H-460 human
lung cancer and MCF-7 human breast cancer cells.
```

Indexing

Indexing in CAplus consists of CAS classification codes, supplementary terms, and indexing terms. The indexing fields can have other names in other bibliographic databases, such as Controlled Term (CT). Consult the relevant Database Summary Sheet for details.

Classification Code, CC

The Classification Code is a broad area or section of chemistry and chemistry-related sciences, as defined by CAS.

```
CC   10-1 (Microbial, Algal, and Fungal Biochemistry)
     Section cross-reference(s): 26
```

Supplementary Terms, ST

Supplementary Terms are keywords that identify main concepts for the original document's subject matter and usually are related directly to terms used by authors in the original publication. These terms can include newly coined scientific or technical terminology and are not standardized.

```
ST   moromycin angucycline anticancer antibiotic natural product Streptomyces;
     lung mammary neoplasm antitumor antibiotic moromycin glycosylangucycline
```

Index Terms, IT

The Index Term field includes standardized, in-depth indexing for the original document's subject matter. Identifying the database's controlled vocabulary is an effective way to assist with retrieving relevant records. The use of index terms is discussed in the *Evaluate and Display Search Results* workbook.

```
IT   Antitumor agents
        (antibiotic; isolation and structure elucidation of
        C-glycosylangucycline-type antibiotics moromycins A and B from
        Streptomyces sp. KY002)
IT   Antibiotics
        (antitumor; isolatio
        C-glycosylangucyclin
        Streptomyces sp. KY0
•••
IT   1044869-32-0P, Moromycin A
     RL: BSU (Biological study, unclassified); NPO (Natural product
     occurrence); PRP (Properties); PUR (Purification or recovery); BIOL
     (Biological study); OCCU (Occurrence); PREP (Preparation)
        (isolation and structure elucidation of C-glycosylangucycline-type
        antibiotics moromycins A and B from Streptomyces sp. KY002)
```

CAS Registry Numbers® and related indexing are explained in the substance searching workbooks.

SEARCH AND DISPLAY FIELDS

The content of a bibliographic record is organized into fields which can be searched and/or displayed. For example, author names are extracted and placed into the author field. Each database summary sheet includes a list, with examples, of all the search and display fields in the database.

The field codes are usually identical; the search field is indicated using a forward slash (/). The SEARCH command statement is entered using the following format: SEARCH <the search term>/ < field code>. For example, to search the term “moromycin” in the title field, you would enter the following command:

=> SEARCH MOROMYCIN/TI

The following table lists search and display fields and field codes commonly found in bibliographic databases.

COMMON SEARCH FIELD	CONTENT	SEARCH FIELD CODE	DISPLAY FIELD CODE
Title	Document title	/TI	TI
Author	Author names	/AU	AU
Corporate Source	Company or organization name for author affiliation	/CS	CS
Publication Year	Publication year of the original publication	/PY	PY
Language	Language of the original publication	/LA	LA
Document Type	Type of document, e.g., journal, patent	/DT	DT

Basic Index, BI

Every STN database includes a special search field called the Basic Index (BI). The Basic Index is composed of a predefined set of fields. If you do not specify a search field, the Basic Index is searched by default.

The fields included in the Basic Index vary between databases. The Database Summary Sheet lists the fields included in the Basic Index for a particular database. It is always the first search field listed.

For example, the Basic Index for CAplus is composed of single words from the Title (TI), Supplementary Terms (ST), Index Terms (IT), and Abstract (AB) fields.

INTRODUCTION TO AN ONLINE SESSION

STN uses a command-line interface, which means that you must type commands instead of either selecting them from a menu or clicking an icon. To conduct an online STN session, you need a minimum of four commands:

- FILE
- SEARCH
- DISPLAY
- LOGOFF

Typing the entire command at the command prompt and pressing **Enter** on the keyboard is referred to as novice mode. When you enter information in this manner, STN prompts you for the additional instructions required to execute the command.

You can also enter commands in a shorter format, referred to as expert mode. When using expert mode, STN assumes that you understand any default values associated with the command and have entered any instructions that differ from the default values. More information about the default values is discussed later in this workbook.

NOVICE MODE	EXPERT MODE	TO
FILE	FIL	Enter the database
SEARCH	S	Conduct your search
DISPLAY	D	View search results
LOGOFF	LOG Y	End the STN session

ONLINE SESSION EXAMPLE

Planning and conducting a bibliographic search involves several steps. The following example shows how to search on a single search term in the CAplus database.

Search Question

What references are available in CAplus concerning the Streptomyces bacteria?

Plan the Search

Plan a search prior to your STN session. The following suggestions can help you prepare:

- Identify the database in which to conduct the search
- Select the search term
- Choose a search field for your search term(s) and the DISPLAY format or field(s) you want to use

While several STN databases contain data about this topic, CAplus is used because it has excellent coverage in the area of biochemistry.

The search question consists of one concept that can be covered by one search term: “Streptomyces.” The word “bacteria” is not used because it is too general. “Streptomyces” is a better choice since it is the genus of bacteria given in the search question.

The Basic Index is chosen because the term can be searched in multiple fields simultaneously.

ACTION	OPTION CHOSEN
Identify database	CAplus
Identify main concepts (search terms)	Streptomyces
Choose search field	Basic Index

Open a Database using the FILE Command

To open a database in STN, type the word FILE at the command prompt. STN then asks you for any instructions it needs to complete the command, in this case, the name of the database. Type the name of the database and then press **Enter**.

When you enter the database, STN displays the time and date of entry, copyright information, and any database-specific information.

```
FILE 'HOME' ENTERED AT 13:37:44 ON 19 MAR 2010

=> FILE CAPLUS
COST IN U.S. DOLLARS

FULL ESTIMATED COST

FILE 'CAPLUS' ENTERED AT 17:44:33 ON 27 MAY 2010
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2010 AMERICAN CHEMICAL SOCIETY (ACS)

Copyright of the articles to which re
held by the publishers
*
*
*
*
FILE COVERS 1907 - 27 May 2010  VOL 152 ISS 22
FILE LAST UPDATED: 26 May 2010  (20100526/ED)
```

You can use both uppercase and lowercase to enter commands.

When you open a database (file), the **file banner** provides up-to-date information about the database.

```
REVISED CLASS FIELDS (/NCL) LAST RELOADED:  Apr 2010
USPTO MANUAL OF CLASSIFICATIONS THESAURUS ISSUE DATE:  Apr 2010

CAplus now includes complete International Patent Classification (IPC)
reclassification data for the second quarter of 2010.

CAS Information Use Policies apply and are available at:
http://www.cas.org/legal/infopolicy.html

This file contains CAS Registry Numbers for easy and accurate
substance identification.
```

Run the SEARCH Command

The SEARCH command retrieves records from the database. After you enter SEARCH, you are prompted for your search term(s).

To search a term in the Basic Index, enter the SEARCH command, the search term, BI as the field code and then press **Enter**. If you do not specify a search field, the Basic Index is assumed.

```
=> SEARCH STREPTOMYCES/BI
L1       40487 STREPTOMYCES/BI
```

After executing the search, STN returns an answer set containing the records that satisfy your search criteria. The answer set is designated with an L-number. In this example, L1 contains 40487 records that have the term “Streptomyces” in the Basic Index.

Records are arranged in reverse chronological order: Answer 1 corresponds to the newest record in the answer set, and Answer 40487 corresponds to the oldest record in the answer set.

View Information using the DISPLAY Command

To view information from the records in an answer set, use the DISPLAY command.

After you enter DISPLAY at the command prompt, STN prompts you for all the information that it needs to execute the command:

- Answer set L-number
- Answer number or range
- Display format or field(s)

To obtain additional information, enter a question mark (?) at any command prompt.

A format refers to a predefined set of fields that are displayed together. For bibliographic databases, BIB is the default display format.

Each database summary sheet lists the display formats and fields available for displaying answers from the database.

To find your search term in the display fields, STN highlights the search term in red. This feature is called hit-term highlighting.

```
=> DISPLAY
ENTER (L1), L# OR ?:L1
ENTER ANSWER NUMBER OR RANGE (1):4
ENTER DISPLAY FORMAT (BIB):BIB

L1   ANSWER 4 OF 40487  CAPLUS  COPYRIGHT 2010 ACS on STN
AN   2010:308384  CAPLUS  Full-text
TI   RNA degradation and the regulation of antibiotic synthesis in Streptomyces
AU   Jones, George H.
CS   Department of Biology, Emory University, Atlanta, GA, 30319, USA
SO   Future Microbiology (2010), 5(3), 419-429
     CODEN: FMUIAR; ISSN: 1746-0913
PB   Future Medicine Ltd.
DT   Journal
LA   English
```

In this example, record number 4 is displayed.

Hit-term highlighting of the search term “Streptomyces.”

Recall that field names can be displayed as complete words or as abbreviated field codes. In the example on page 5, the Indented Bibliographic (IBIB) format is used to display the complete name of the fields and associated information.

Sub-Command Prompts and Defaults for the DISPLAY Command

After you enter the DISPLAY command at the command prompt, STN asks for the additional information it needs to execute the command. These questions, also known as sub-command prompts or colon prompts, end with a colon (:). At this prompt, you must enter a response before STN can return you to the primary command prompt (=>).

If you accidentally go to a sub-command prompt, you can return to the primary command prompt by typing **END**.

Each sub-command prompt for the DISPLAY command also includes a default option, enclosed in parentheses. To select the default option in response to a prompt, type a period (**.**) and then press **Enter**.

DISPLAY PROMPT FOR	DEFAULT SETTING
Answer set L-number	Last L-number created
Answer number or range	Answer 1 (the most recent)
Display format	BIB (bibliographic information)

Iterative Steps

After reviewing records from your initial search, identify additional terms, such as synonyms, that you want to add to your search. Revising, searching, and displaying answers is often an iterative process.

Other workbooks, including *Search for Multiple Terms in Bibliographic Databases*, *Refine a Search*, *Broaden a Search*, and *Evaluate and Display Search Results* provide additional information about these techniques.

End the Session using the LOGOFF Command

The LOGOFF command is used to end your STN session. When you type LOGOFF at the command prompt, STN displays a prompt with the following options.

LOGOFF OPTION	RESULT
Y	All L-numbers from your session are deleted and your session is ended.
N	Logoff is cancelled and your online session remains active.
HOLD	Your online session is suspended, at no charge, for up to 120 minutes. Log in to STN to resume your session.

LOG Y is the expert way to end your STN session.

When you end a session, STN displays the session summary and cost information. This summary is stored in the transcript of your session if you chose to save your transcript file. For more information about transcripts, see the *Work with Transcripts* workbook.

```
=> LOGOFF
ALL L# QUERIES AND ANSWER SETS ARE DELETED AT LOGOFF
LOGOFF? (Y)/N/HOLD:Y
```

Resume a Session after a LOGOFF HOLD

When you return to STN after you enter LOGOFF HOLD, type DISPLAY HISTORY (D HIS) at the command prompt to see a list of your L-numbers and the search associated with each one. These L-numbers are active and you can use them to continue your online session.

SUMMARY

- The STN Database Summary Sheets provide information about content and features of databases
- Planning and conducting a search involves the use of commands and the instructions used by the commands to carry out the task
- FILE, SEARCH, DISPLAY, and LOGOFF are the four basic commands

Suggested Search Strategy

STEP	ACTION	EXAMPLE
1	Plan the search	• Identify the database in which to conduct the search • Select the search terms • Choose a field to search • Choose the DISPLAY format or field(s) you want to use (discussed in the *Evaluate and Display Search Results* workbook)
2	Open a database with the FILE command	=> FILE CAPLUS
3	Run the SEARCH command	=> SEARCH STREPTOMYCES/BI
4	Display information with the DISPLAY command	=> DISPLAY ENTER (L1), L# OR ?:L1 ENTER ANSWER NUMBER OR RANGE (1):1, 40590 ENTER DISPLAY FORMAT (BIB): BIB
5	End the session with the LOGOFF command	=> LOGOFF YES

ADDITIONAL RESOURCES

RESOURCE	LOCATION	USED FOR
CAS Learning Solutions	learningsolutions.cas.org	Self-directed learning resources and instructor-led training events
STN Training Resources	www.cas.org	• Recorded e-seminars • Database Summary Sheets • User documentation
STN Customer Centers	CAS: help@cas.org FIZ: helpdesk@fiz-karlsruhe.de JAICI: customer@jaici.or.jp	Assistance with STN searches and account management

PRACTICE EXERCISE

Practice Exercise

Find references for chocolate in the CAplus database*. Display the newest and the oldest reference.

HINT: Remember that answers are deposited into an L-number in reverse chronological order.

*You could also conduct this search in the LCA learning database, the learning version of CAplus. Since it is a much smaller, static database, your answer set will be much smaller than if you conduct the search in CAplus.

CONCEPT 1	CONCEPT 2	CONCEPT 3
Chocolate		

SUGGESTED SOLUTION

These search strategy steps demonstrate the techniques and tools described in this workbook. Other approaches are possible and any search strategy is always best designed based on the needs of the searcher. In addition, databases are frequently updated with new records; therefore, replicating these exercises can result in an answer set with numbers that differ from what is shown.

Suggested Search Strategy

STEP	ACTION	EXAMPLE
1	Plan the search	• The search will be conducted in CAplus • Chocolate is the search term and it will be searched in the basic index • The BIB display format will be used
2	Open a database	=> FILE CAPLUS
3	Conduct the search	=> SEARCH CHOCOLATE/BI

STEP	ACTION	EXAMPLE
4	Display the data of interest	=> DISPLAY ENTER (L1), L# OR ?:L1 ENTER ANSWER NUMBER OR RANGE (1):1, 10189 ENTER DISPLAY FORMAT (BIB):. (Remember that entering a period tells STN to accept the default, which is BIB in this example.)
5	End the session	=> LOGOFF Y

Transcript Highlights

```
=> FILE CAPLUS

=> S CHOCOLATE
L2       10189 CHOCOLATE
                (CHOCOLATE OR CHOCOLATES)

=> D 1, 10189

L2   ANSWER 1 OF 10189  CAPLUS  COPYRIGHT 2010 ACS on STN
AN   2010:638019  CAPLUS  Full-text
TI   Detection of psilocybin mushroom analogs in chocolate: incorporating
     current events into the undergraduate teaching laboratory
AU   Huskins, Brandon; Dockery, Christopher R.
CS   Department of Chemistry and Biochemistry, Kennesaw State University,
     Kennesaw, GA, 30144, USA
SO   Chemical Educator (2009), 14(6), 236-238
     CODEN: CHEDF5; ISSN: 1430-4171
PB   Chemical Educator
DT   Journal; (online computer file)
LA   English

L2   ANSWER 10189 OF 10189  CAPLUS  COPYRIGH
AN   1906:1531  CAPLUS  Full-text
DN   0:1531
TI   Refrigeration, cold storage, and ice-ma
AU   Wallis-Taylor, A. J.
SO   Publisher: (Lockwood & Son, London).  $4.50 net.
     Reviewed in: J. Am. Chem. Soc. 24(8) 781-783, 1902
DT   Book
LA   Unavailable

=> LOG Y
```

Display record 1 to see the newest reference and record 10189 for the oldest reference in the answer set. Records in the answer set are in reverse chronological order.

The Full-text link takes you to ChemPort®, which gives you options available for viewing the full text. The options range from publisher web sites to ordering the document.

In this transcript, the language is listed as Unavailable because this record is from 1906. The language field was added to CAplus in 1967.