SEARCH AND DISPLAY SUBSTANCE INFORMATION

COPYRIGHT© 2010 AMERICAN CHEMICAL SOCIETY

ALL RIGHTS RESERVED; PRINTED IN THE USA.
Quoting or copying of material from this publication for educational purposes is encouraged, provided that CAS is acknowledged as the source of the material.

WHAT YOU WILL LEARN

In this workbook, you will learn how to:

- Use the SEARCH command in the following CAS REGISTRY℠ fields:
  - Chemical Name (CN)
  - Basic Index (BI)
- View content that is included when using the DISPLAY SCAN format
- View specific parts of a REGISTRY record

RECOMMENDED SUBSTANCE LEARNING PATH

The *Substance Search Learning Path* is a series of workbooks designed to teach skills required to perform substance searching using STN®. The content assumes a working knowledge of basic STN search techniques. For information about the fundamentals of STN online searching, refer to the *Basic Text Search Learning Path*.

If you are new to substance searching, it is recommended that you use these workbooks in the order shown, since each one builds on concepts and skills covered in preceding workbooks. If you are an experienced user, the *Substance Search Learning Path* can be used for reviewing specific techniques or acquiring additional skills.

1. Understand Bibliographic Versus Substance Databases
2. **Search and Display Substance Information**
3. Retrieve CAplus℠ References Associated with Substances
4. Retrieve Bibliographic References from Non-CAS Databases for Substances
5. Search Molecular Formulas
6. Find Property Data in CAS REGISTRY

Workbook Conventions

- STN Express® is used to illustrate concepts throughout these training workbooks, although the command line techniques also apply to STN® on the Web℠
- The REGISTRY database is used throughout the *Substance Learning Path*

SEARCH CHEMICAL NAMES IN REGISTRY

The *Understand Bibliographic Versus Substance Databases* workbook describes how you can find a chemical substance using the chemical name, molecular formula, property data, or structures. This workbook focuses on several techniques used for conducting chemical name searches.

USE THE EXPAND COMMAND FIRST

To search for a specific chemical name in the CN field, the chemical name must be spelled exactly as it appears in the database. Start by using the EXPAND command to ensure that the chemical name appears in the database prior to conducting the search as it is easy to mistype chemical names.

Find Simple Chemical Names

A chemical name search using the CN field works effectively when the name is:

- A trade name, such as Ambien®
- An easily spelled common name, such as acetone
- A simple International Union of Pure and Applied Chemistry (IUPAC) name, such as 2-Propanone

CAS analysts add chemical names to REGISTRY for substances found in documents, such as patents and journals. New names can be added to existing records, and new REGISTRY records are created for substances not previously encountered in published literature. REGISTRY is a robust source of chemical names with thousands of journals and patents from more than 63 patenting authorities used as source documents.

Simple Trade Name Example

```
=> FILE REG

=> E AMBIEN/CN
E1            1     AMBICROMIL/CN
E2            1     AMBICURE XP 150/CN
E3            1 --> AMBIEN/CN
E4            1     AMBIENT AIR/CN
E5            1     AMBIFARIN A/CN
E6            1     AMBIFIX BLACK BFGR/CN
E7            1     AMBIFIX NAVY HER/CN
E8            1     AMBIFIX YELLOW VRNL/CN
E9            1     AMBIGOL A/CN
E10           1     AMBIGOL B/CN
E11           1     AMBIGOL C/CN
E12           1     AMBIGUINE/CN

=> S E3
L1            1 AMBIEN/CN

=> D
```

```
L1   ANSWER 1 OF 1  REGISTRY  COPYRIGHT 2010 ACS on STN
RN   99294-93-6  REGISTRY
ED   Entered STN:  30 Nov 1985
CN   Imidazo[1,2-a]pyridine-3-acetamide, N,N,6-trimethyl-2-(4-methylphenyl)-,
     (2R,3R)-2,3-dihydroxybutanedioate (2:1)  (CA INDEX NAME)
OTHER CA INDEX NAMES:
CN   Imidazo[1,2-a]pyridine-3-acetamide, N,N,6-trimethyl-2-(4-methylphenyl)-,
     [R-(R*,R*)]-2,3-dihydroxybutanedioate (2:1)
OTHER NAMES:
CN   Ambien
CN   Ivadal
CN   Niotal
CN   SL 800750-23N
CN   Stilnoct
CN   Stilnox
CN   Zolpidem hemitartrate
CN   Zolpidem tartrate
*
*
*
```

Although Ambien is easy to spell, the CA Index Name is quite complex.

Find Complex Chemical Names

Names that describe the chemistry of a substance can be complex, as seen in the previous example for Ambien. These chemically significant names are created using nomenclature rules to describe the substance. The IUPAC nomenclature system is widely used and taught in chemistry classes. CAS also has created a nomenclature system that is even more descriptive than the IUPAC system. It is used to create the CA Index Name.

Attempting to search for a CA Index Name can be very challenging because every character (comma, dash, space, and parenthesis) must be typed exactly as it appears in the database entry. Therefore, when looking for substances that do not have simple trade or common names, use the Basic Index or a structure search.

Search Chemical Name Fragments in the Basic Index

The Basic Index field (BI) in REGISTRY includes chemical name fragments, molecular formula fragments, and Collective Index codes.

Chemical name fragments include:

- Basic segments of the full name – the smallest chemically distinct name fragments remaining after punctuation are removed (i.e., meth and oxy)
- Recombined segments – basic segments recombined (i.e., methoxy)

Generally, name fragment searching in the Basic Index is used as a refinement strategy when conducting structure and formula searches. However, if you have the correct nomenclature for a substance, you can often retrieve the record by conducting a chemical name fragment search.

First, use the EXPAND command on the chemical name fragments that you are interested in retrieving. Then, construct the search query by connecting each search term with the (L) or (XA) proximity operator. These operators will require each fragment to appear in a single chemical name. The (XA) operator allows for multiple occurrences of the same word in a name. The following example uses fragments from CA Index Name for Ambien.

Chemical Name Fragment Search in the Basic Index

```
=> FILE REG

=> E IMIDAZO 5
E1            2     IMIDAZINE/BI
E2            2     IMIDAZLE/BI
E3       621403 --> IMIDAZO/BI
E4            4     IMIDAZOANTHRA/BI
E5            1     IMIDAZOBENZENE/BI

=> E N,N,6 5
E6            1     N,N,5B,15B/BI
E7            1     N,N,5B,17B/BI
E8         2284 --> N,N,6/BI
E9           22     N,N,6'/BI
E10           1     N,N,6''/BI

=> E METHYLPHENYL 5
E11           1     METHYLPHENTR/BI
E12           1     METHYLPHENTRAMINE/BI
E13     3790937 --> METHYLPHENYL/BI
E14         148     METHYLPHENYLACET/BI
E15           5     METHYLPHENYLACETALDEHYDE/BI

=> E PYRIDINE 5
E16           9     PYRIDINDOLE/BI
E17           5     PYRIDINDOLOL/BI
E18     1887807 --> PYRIDINE/BI
E19           2     PYRIDINE-KN/BI
E20       41219     PYRIDINEACET/BI

=> E ACETAMIDE 5
E21           2     ACETAMIDATOTRI/BI
E22           2     ACETAMIDATOTRIS/BI
E23     2761093 --> ACETAMIDE/BI
E24           1     ACETAMIDEBENZ/BI
E25           1     ACETAMIDEBENZALDEHYDE/BI

=> S E3 (L) E8 (L) E13 (L) E18 (L) E23
L1           96 IMIDAZO/BI (L) "N,N,6"/BI (L) METHYLPHENYL/BI (L) PYRIDINE/BI
                (L) ACETAMIDE/BI

=> E DIHYDROXY 5
E26           1     DIHYDROXPROPOXY/BI
E27           3     DIHYDROXUXUARIN/BI
E28      506891 --> DIHYDROXY/BI
E29           1     DIHYDROXYA/BI
E30          11     DIHYDROXYABIET/BI

=> S L1 (L) E28
L2           16 L1 (L) DIHYDROXY/BI

=> D IN STR 1-16

*
*
*

L2   ANSWER 16 OF 16  REGISTRY  COPYRIGHT 2010 ACS on STN
IN   Imidazo[1,2-a]pyridine-3-acetamide,
     N,N,6-trimethyl-2-(4-methylphenyl)-, (2R,3R)-2,3-dihydroxybutanedioate
     (2:1)
```

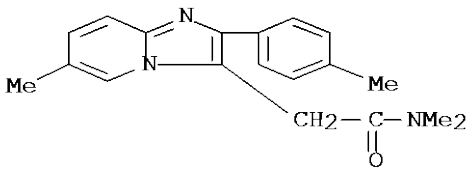
Using the EXPAND command, you can display 5-25 E-numbers. In this example, five E-numbers are requested.

Many of the chemical name fragments are associated with a large number of records.

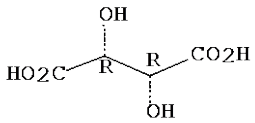
96 records are more than you want to display, so use an additional term to narrow the search.

Display the Index Name (IN) and Structure (STR) for each record and then isolate the record of interest: Answer 16.

```
CM   1
```




```
**PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT**

     CM   2

Absolute stereochemistry.
```




```
**PROPERTY DATA AVAILABLE IN THE

=> D 16
*
*
*
```

Answer number 16 is the substance of interest. Type D 16 at the command prompt to display the default IDE format.

Additional information about searching chemical name fragments is available in intermediate and advanced training documents.

USEFUL DISPLAY FORMATS IN REGISTRY

This section highlights frequently used display formats in REGISTRY. Consult the REGISTRY Database Summary Sheet (DBSS) for a complete list of display formats.

DISPLAY INDIVIDUAL FIELDS

You can display many of the individual fields that are part of a REGISTRY record. The following table lists several useful display codes for individual fields.

FIELD NAME	DISPLAY FIELD OPTIONS
Registry Number	RN
Chemical Name	• CN (up to 50 names) • FCN (The Full Chemical Names format lists every chemical name in the record) • IN for just the CA Index Name*
Molecular Formula	MF
Entry Date	ED
Source of Registration	SR
Structure	STR

* IN is a display only field; it is not searchable

DISPLAY FORMATS

Display formats are predefined groups of field codes that can be displayed. The default display format, IDE, is an example. The following table lists several frequently used display formats and the displayed content.

DISPLAY FORMAT	DISPLAYED CONTENT
SCAN	Index Name, Sequence Length, Molecular Formula, Substance Class Identifier, Structure, and Composition (for alloys) SCAN is a random display without answer numbers.
FIDE	All substance information including all names and Ring System Data (RSD)*, CAplus super roles* and document types, and all property tables (EPROP, ETAG, PPROP)*

DISPLAY FORMAT	DISPLAYED CONTENT
IDE	Registry Number, Entry Date, Chemical Names, File Segment, Molecular Formula, Class Identifier, Source, Registry Number Locator, Structure Diagram, and number of references in CA[SM] and CAplus
SQIDE	Everything in the IDE format, as well as several fields associated with sequence data

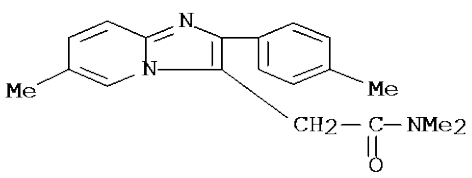
* Use of Ring System Data and CAplus super roles are taught in the intermediate and advanced materials. Property searching is covered in the *Find Property Data in CAS REGISTRY* workbook.

Many formats and display codes are also available for property data and are discussed in the *Find Property Data in REGISTRY* workbook.

Example of the FIDE Format for Ambien

The first part of the record is the IDE data (not shown here).

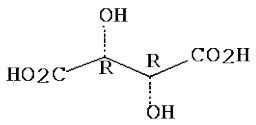
```
*
*
*
DT.CA  CAplus document type:  Journal; Patent
RL.P   Roles from patents:  ANST (Analytical study); BIOL (Biological study);
       PREP (Preparation); PROC (Process); PRP (Properties); PRPH (Prophetic);
       RACT (Reactant or reagent); USES (Uses)
RL.NP  Roles from non-patents:  ANST (Analytical study); BIOL (Biological
       study); FORM (Formation, nonpreparative); PREP (Preparation); PROC
       (Process); PRP (Properties); USES (Uses)

Ring System Data

Elemental|Elemental| Size of |Ring System|   Ring    |   RID
Analysis |Sequence |the Rings|  Formula  |Identifier|Occurrence
   EA    |   ES    |   SZ    |    RF     |   RID    |  Count
=========+=========+=========+===========+==========+==========
C6       |C6       |6        |C6         |46.150.18 | 1 in CM
         |         |         |           |          |1
C3N2-C5N |NCNC2-NC5|5-6      |C7N2       |333.868.10| 1 in CM
         |         |         |           |          |1



     CM   1

     CRN  82626-48-0
     CMF  C19 H21 N3 O
```



CM 2

CRN 87-69-4
CMF C4 H6 O6

Absolute stereochemistry.



Experimental Properties (EPROP)

PROPERTY (CODE)	VALUE	NOTE
Melting Point (MP)	196 deg C	(1) NLM

(1) "Hazardous Substances Data Bank" data were obtained from the National Library of Medicine (US)

Experimental Property Tags (ETAG)

PROPERTY	NOTE
Compressibility	(1) CAS
Crystal Structure	(2) CAS
IR Absorption Spectra	(3) CAS
Particle Size	(1) CAS
Solubility	(4) CAS
Specific Surface Area	(1) CAS
X-Ray Diffraction Pattern	(2) CAS

(1) Emshanova, S. V.; Pharmaceutical Chemistry Journal 2007 V41(12) P659-661 CAPLUS
(2) Halasz, Ivan; Journal of Pharmaceutical Sciences 2010 V99(2) P871-878 CAPLUS
(3) El Zeany, B. A.; Journal of Pharmaceutical and Biomedical Analysis 2003 V33(3) P393-401 CAPLUS
(4) Hays, Patrick A.; Journal of Forensic Sciences 2005 V50(6) P1342-1360 CAPLUS

See HELP PROPERTIES for information about property data sources in REGISTRY.
200 REFERENCES IN FILE CA (1907 TO DATE)
202 REFERENCES IN FILE CAPLUS (1907 TO DATE)

SUMMARY

- Simple chemical names can be searched directly in the CN field
- It is recommended that you conduct a structure search for substances with complex chemical names
  - Given an accurate complex name, you can search the chemical name fragments in the Basic Index, connecting the fragments with the (L) or (XA) proximity operator
- In REGISTRY, the DISPLAY SCAN format includes the CA Index Name, Molecular Formula, Collective Index codes, and the Structure (when available)
- Many display field codes and display formats are available in REGISTRY; complete details can be found in the STN Database Summary Sheets

SEARCH STRATEGY BEST PRACTICES

These search strategy steps showcase the techniques and tools explored in this workbook. Other approaches are possible and any search strategy is always best determined by the intentions and needs of the searcher.

STEP	ACTION	EXAMPLE
1	Open the database	=> FILE REG
2	Confirm that each search term is in the database	=> E AMBIEN/CN
3	Conduct the search (alternatively, search the correct E-number)	=> S AMBIEN/CN
4	Display the record in the default format	=> D

PRACTICE EXERCISES

Practice Exercise 1

a. Retrieve the REGISTRY record for ibuprofen.

b. Display the default identification format.

c. Notice the STN message at the end of the CN field: “ADDITIONAL NAMES NOT AVAILABLE IN THIS FORMAT - Use FCN, FIDE, or ALL for DISPLAY.” Display the FCN format.

*You could also conduct this search in the learning LREGISTRY[SM] database. Since it is a much smaller, static database, your answer set will be much smaller than if you conduct the search in REGISTRY.

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 results that differ from what is shown.

Suggested Search Strategy

STEP	ACTION	EXAMPLE
1	Open the database	=> FILE REG
2	Verify that each search term is in the database	=> E IBUPROFEN/CN
3	Conduct the search	=> S E3
4	Display the record	=> D
5	Display all of the chemical names	=> D FCN

Transcript Highlights

```
*
*
*
CN   Ibuprofen
ADDITIONAL NAMES NOT AVAILABLE IN THIS FORMAT - Use FCN, FIDE, or ALL for
     DISPLAY
DR   58560-75-1, 139466-08-3
MF   C13 H18 O2
*
*
*
```

Practice Exercise 2

This exercise provides the opportunity to learn more about the numerous sources of data used to create the largest substance database in the world.

a. EXPAND on the letter L in the Source of Registration field and show 25 E-numbers (i.e., E L/SR 25). EXPAND a second time, showing 25 more E-numbers.

b. Retrieve the Lipid Maps Structure Database and search that E-number.

c. Refine the answer set by searching on the term Hexacosenamide in the Basic Index.

d. Display the Registry Numbers for all answers.

e. Display the FIDE format for Registry Number 958763-62-7.

*You could also conduct this search in the learning LREGISTRY[SM] database. Since it is a much smaller, static database, your answer set will be much smaller than if you conduct the search in REGISTRY.

Suggested Search Strategy

STEP	ACTION	EXAMPLE
1	Open the database	=> FILE REG
2	EXPAND in the Source of Registration field	=> E L/SR 25
3	EXPAND until you find the Lipid Maps Structure Database	=> E 25
4	Refine the answer set by searching on Hexacosenamide	=> S L1 AND HEXACOSENAMIDE
5	Display the Registry Numbers for each answer	=> D 1-2 RN
6	Display the FIDE format for Registry Number 958763-62-7	=> D FIDE 1

Transcript Highlights

```
*
*
*
=> E 25
E38         2198     LIPID MAPS CONSORTIUM/SR
*
*
*
=> S E38
L1          2198 "LIPID MAPS STRUCTURE DATABASE (LIPID MAPS CONSORTIUM)"/SR
*
*
```

```
*
=> S L1 AND HEXACOSENAMIDE
             52 HEXACOSENAMIDE
L2            2 L2 AND HEXACOSENAMIDE

=> D 1-2 RN

L2   ANSWER 1 OF 2  REGISTRY  COPYRIGHT 2010 ACS on STN
RN   958763-62-7  REGISTRY

L2   ANSWER 2 OF 2  REGISTRY  COPYRIGHT 2010 ACS on STN
RN   958760-66-2  REGISTRY

=> D FIDE
*
*
*
```

ADDITIONAL RESOURCES

RESOURCE	LOCATION	USED FOR
STN Training Resources	www.cas.org	• Recorded e-seminars • STN 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