Overview of Structural Databases

CAS is a division of the American Chemical Society

Information is the foundation for making important business-critical decisions

"Information technology and business are becoming inextricably interwoven. I don't think anybody can talk meaningfully about one without talking about the other."

Bill Gates
Chairman, Board of Directors
Microsoft

Lack of organized data and accurate search systems can be expensive

Finding critical information is a costly, global problem

Source: Information World Review
www.iwr.co.uk

You can start with Google for background, but…

Relying on information from the Internet is like playing roulette: hit or miss!

- It is not always
  - Complete
  - Documented
  - Accurate
  - Precise

Multi-million dollar business decisions can hinge on one missed patent or even one overlooked molecule

Critical decisions require reliable information

STN is an online database service providing global access to the published scientific literature:

- Journal articles
- Patents
- Chemical structures
- Sequences
- Properties
- Electronic full text

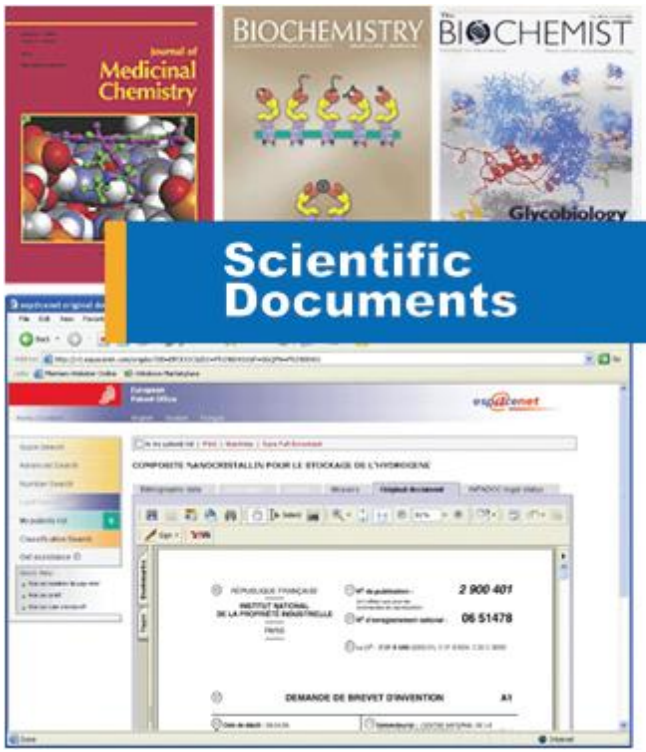
The advantages of using STN are numerous

- Search many high-quality databases at the same time
- Security – STN offers secure Internet access
- Trusted globally by major patent offices
- Access to the latest research – major databases updated daily
- The benefit of an international partnership
- Highly renowned customer support

STN offers an unsurpassed collection of essential scientific and technical databases

- STN databases are uniquely integrated so you can consult several at the same time with one query
  - High-quality, standardized databases in one place
  - Unified command language
  - Value-added indexing by scientists
  - Duplicate answers are easily removed from a multi-database search

STN includes the comprehensive chemistry databases from CAS



Scientific Documents

CAS®

A division of the American Chemical Society

CAS Scientists

High-quality indexing

Basic Searching Commands

- There are only five basic commands that you need to learn to begin searching on STN

USE THIS COMMAND…	WHEN YOU WANT TO…
FILE (FIL)	Enter a single file or multiple files in order to conduct a search
EXPAND (E)	View potential search terms in an easy to browse, alphabetical listing
SEARCH (S)	Perform a search pertaining to a topic of interest
DISPLAY (D)	Look at the record set results in a specified format
LOGOFF (LOG)	Terminate an STN online session

STN provides many other commands so that you can access specialized functionality, but these five commands will go a long way in your searching.

Basic Command Format

- The general format for entering a command on STN is:

```
=> COMMAND Instructions <Enter>
```

Example:

```
=> SEARCH (chocolate or cocoa) <Enter>
```

- For additional information about using STN commands, visit www.cas.org/support/stngen/stndoc/commands.html

Conduct a basic keyword search

Step 1	Understand search question and identify potential keywords
Step 2	Identify relevant database(s)
Step 3	Build a search query
Step 4	Conduct a preliminary search
Step 5	Evaluate answers
Step 6	Modify the search strategy
Step 7	Display answer(s)

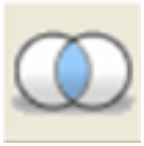
Step 1: Understand search question and identify potential keywords

- Keyword searching is used to build concept-based search queries and is commonly done as free-text searching in the Basic Index
- Have an organized search plan
  - Identify key words
    - Consider synonyms
    - Consider suffixes on root words
    - Don’t forget acronyms and abbreviations
  - Consider how search terms will relate to one another
    - Boolean logic
    - Proximity of search terms

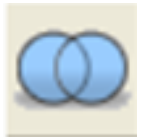
STN Truncation Symbols

SYMBOL	FUNCTION	EXAMPLES	RETRIEVAL POSSIBILITIES
?	Any number of characters (including zero) at the beginning or at the end of a term	BACTERICID?	BACTERICIDE BACTERICIDAL
	Left*-or Right-hand truncation	?ICID?	BACTERICIDE
#	Zero or one character at the end of a term	BACTERICIDE#	BACTERICIDE BACTERICIDES
!	Exactly one character within or at the end of a term	T!!TH	TEETH TOOTH TRUTH
		AMIN!	AMINE AMINO
		ORGANI!ATION	ORGANIZATION ORGANISATION
!!#	Multiple uses of the symbols # and ! are allowed	T!!TH#	TEETH TOOTH TRUTHS

Boolean Logic Operators

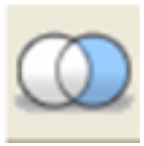


- **"AND"** represents records that mention all of the concepts, anywhere in the record
  - One concept could be in the title, and another could be in the abstract or indexing



- **"OR"** represents records with any of the concepts and synonyms
  - Don't forget to use parentheses with concepts that are OR'ed together

```
=> S (bovine OR cow OR cattle)
```

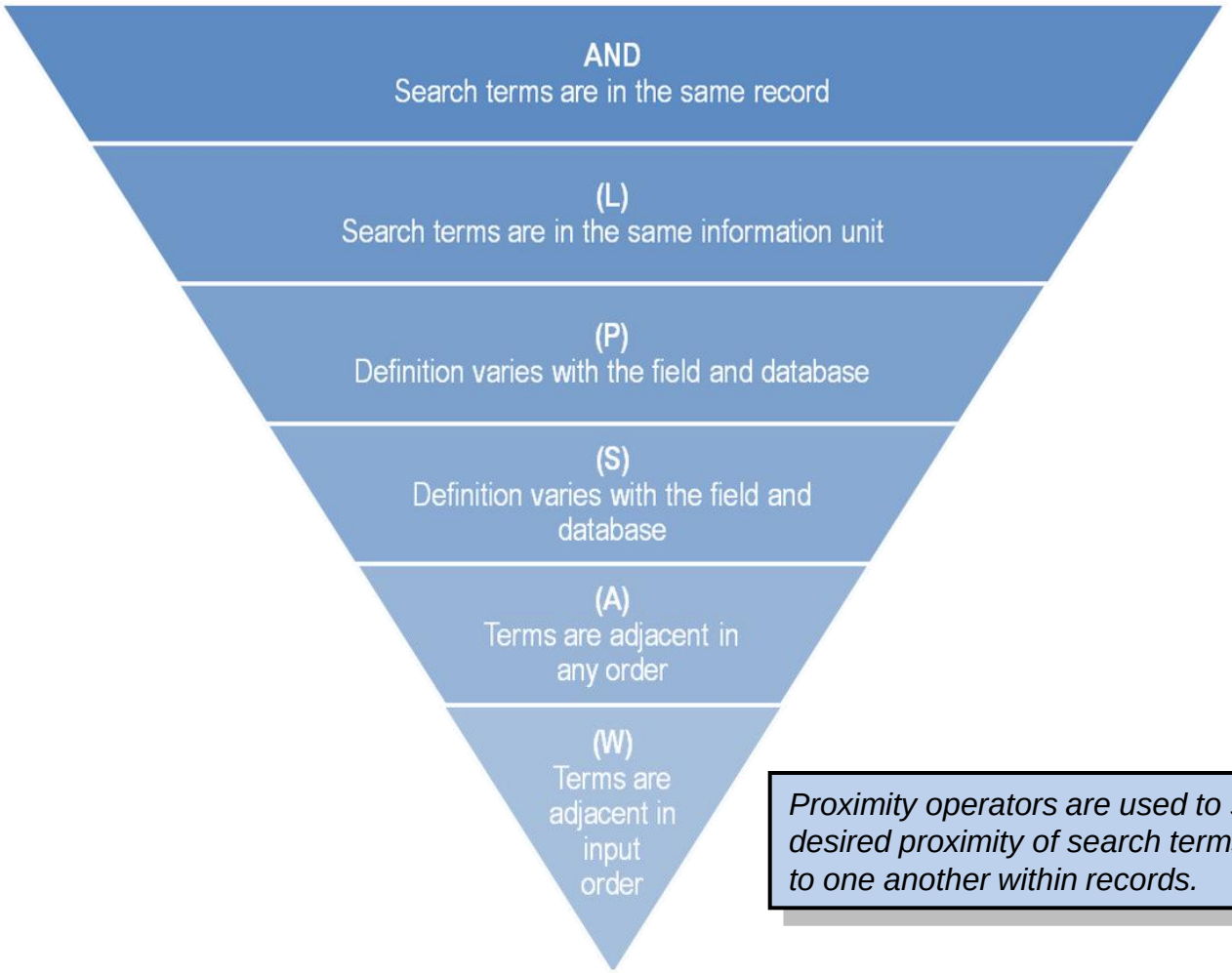


- **"NOT"** removes a concept or answer set from results
  - Use NOT to compare results between sets

```
=> S L3 NOT L1
```

    - Beware of using NOT to remove concepts as you can accidentally remove good results as well

Proximity Operators



Proximity operators are used to specify the desired proximity of search terms with respect to one another within records.

Step 2: Identify relevant database(s)

- A variety of worldwide databases of scientific and technical information are available on STN
- Information about the databases can be found in the following resources:
  - STN Database Guide
    http://stnguide.cas.org/
  - STN Database Summary Sheets
    www.cas.org/support/stngen/dbss/
  - STNGUIDE, a no-cost database that provides searchable access to all of the information covered in the STN DBSS
  - INDEX command
  - CAS Help Desk
    1-800-753-4227 (North America)
- Use the FILE command to enter a database

```
=> FILE CAPLUS
```

Step 3: Build a search query

- Verify search terms using EXPAND
  - The EXPAND command (E) is used to verify that a term of interest is in the database
  - EXPAND results are in an alphanumeric list of indexed terms
  - EXPAND is useful in keyword searching to
    - Determine if a term is in the database
    - Identify related terms of interest
  - EXPAND is a cost-free command

Step 4: Conduct a preliminary search

- Additional terms from the EXPAND search can be found using the CA Lexicon and other thesauri.
  - A Wizard is available to help with this in STN Express®
  - Use of thesauri is covered more fully in advanced STN workshops

Step 5: Evaluate answers

- Once the search has been conducted, review the records to determine if the query retrieved the desired results
- No-cost display formats are useful in keyword searching to identify additional, database-specific terminology to enhance results

Step 7: Display records

- Answers may be displayed in pre-defined formats or custom field displays
- DISPLAY command requires three pieces of information
  - Answer set L-number
  - Format
  - Answer number(s) to be displayed

```
=> DISPLAY L2 IBIB ABS 1 200
```

Why Do a Structure Search?

Structure searching has distinct advantages

- Every chemical compound has a unique structure
- Locate Enantiomeric forms and isotopically labeled molecules of a base structure
- Locate Biologically active salt forms
- Broadly related groups of substances can contain a similar fragment
- Biological activity may be related to a particular substructure
- Polymeric materials will contain similar monomeric units

Locating substance information

There are a number of ways to locate information about a chemical compound

- Chemical name
- Molecular formula
- Structure

Structure Searchable Databases on STN

- A sample of structure searchable databases on STN include

This structure-searchable database	Is useful when you have …
REGISTRY	Organics, inorganics, biochemicals or sequences
BEILSTEIN	Organic substances
GMELIN97	Inorganic substances
WPINDEX, WPIDS, WPIX	Selected structures from Sections B, C, and E of Derwent's World Patent Index (since 8/99)
MARPAT	Markush structures from patent claims
CASREACT, CHEMINFORMRX, DJSMONLINE, DRUGU, PS	Organic reactions

Basic Strategy for Structure Searching

Locating substance information via structure searching is accomplished in 3 steps:

1. Draw the structure using
   - STN Express
   - STN on the Web

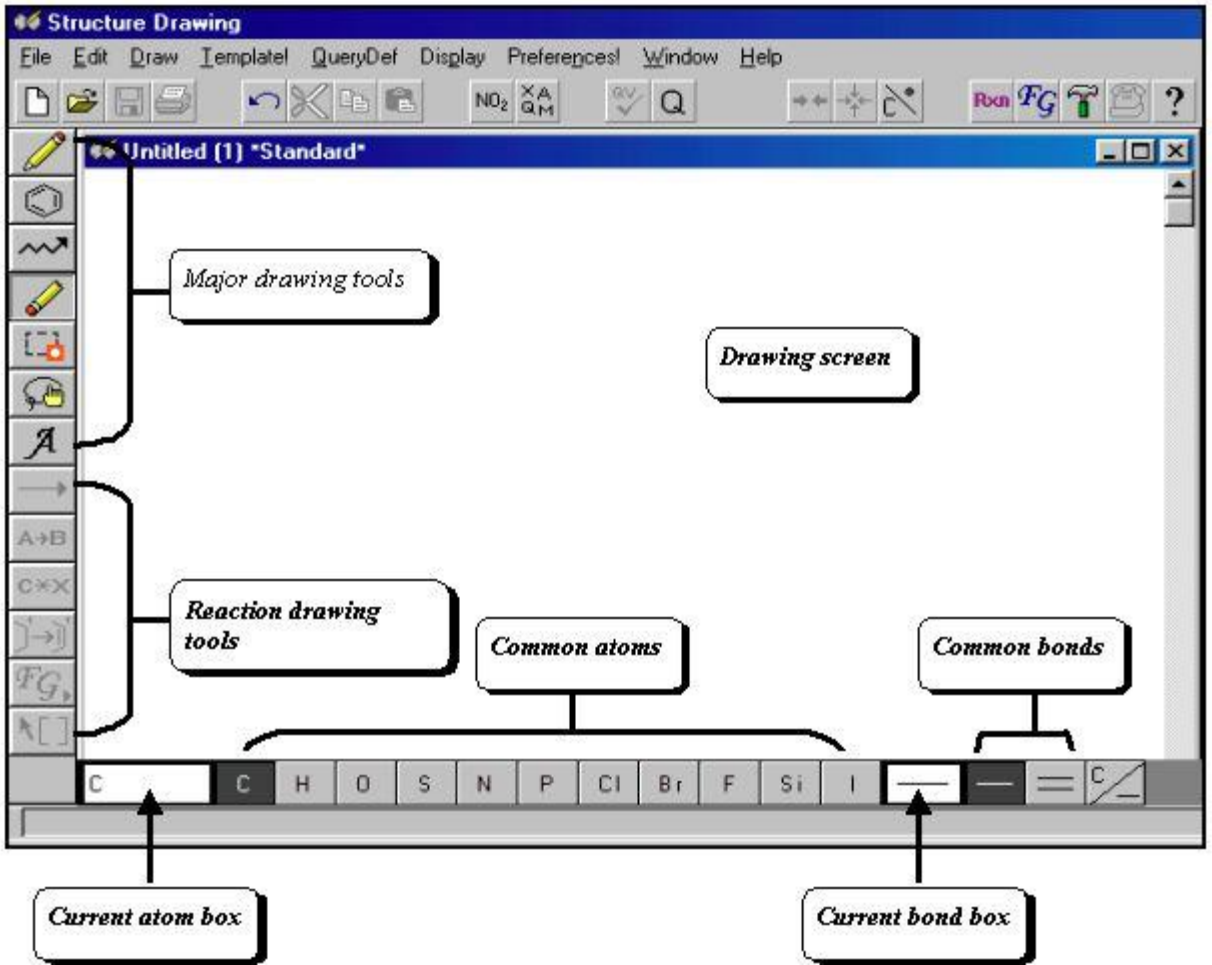
2. Run the structure search in REGISTRY
   - Display structure matches to evaluate results
   - Revise structure, if needed

3. Locate references discussing the structures of interest

Overview

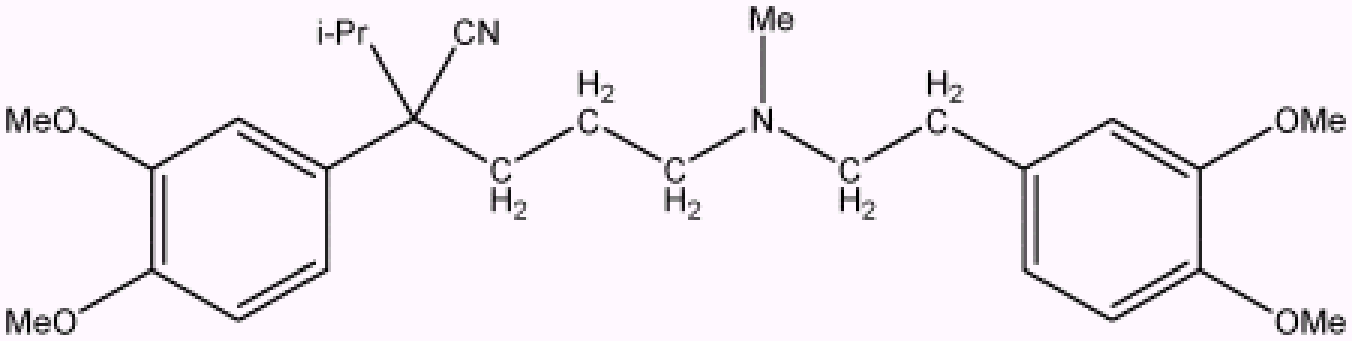
Structure Drawing Screen Components

- The structure drawing screen has 7 main parts:



Structure Searching for a Family of Structures

Search Question: *Locate references discussing the following substance, its salts, and any mixtures containing the substance:*



Run a Sample Structure Search

- A sample (SAM) structure search searches a portion of the database
- It is a no-cost option that is used to evaluate the effectiveness of a structure search by
  - Testing the structure search to ensure it will run within system limits
  - Verifying that the types of answers retrieved are the types of answers desired

Types of structure search

To retrieve this type of substance match	Use this type of structure search:		
	Exact EXA	Family FAM	Substructure SSS
Exact substance	✓	✓	✓
Stereoisomers	✓	✓	✓
Isotopically labeled	✓	✓	✓
Salts		✓	✓
Mixtures		✓	✓
Substitution at open sites			✓

Testing the structure search

. Scope of the structure search

Sample ***(SAM)*** searches a fixed portion of the database to test the query

Full (***FULL)*** searches the full database

```
=> S L1 FAM SAM
SAMPLE SEARCH INITIATED 11:45:03
SAMPLE SCREEN SEARCH COMPLETED -

100.0% PROCESSED      22 ITERATIONS                    9 ANSWERS
SEARCH TIME: 00.00.01
FULL FILE PROJECTIONS:   ONLINE   **COMPLETE**
                         BATCH    **COMPLETE**
PROJECTED ITERATIONS:              159 TO      721
PROJECTED ANSWERS:                   9 TO      360

L2             9 SEA FAM SAM L1
```

This search is projected to run to completion online within system limits.

Structure matches ("answers") are placed in an answer set that is given the next available L-number.

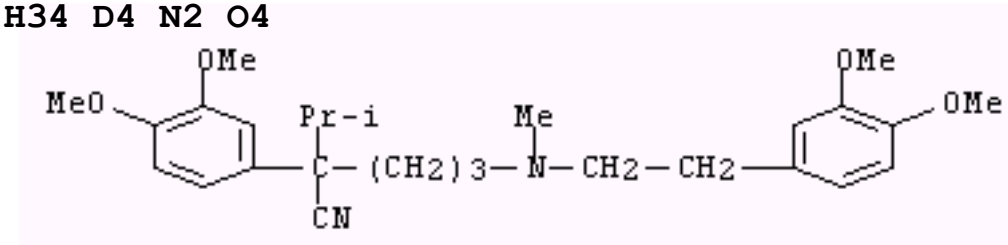
Verifying the answers

- The no-cost D SCAN command is used to evaluate structure search results
- D SCAN randomly selects an answer from the answer set and displays the CAS Index Name, molecular formula and structure

Verifying the answers

```
=> D SCAN

L2    9 ANSWERS   REGISTRY  COPYRIGHT 2006 ACS on STN
IN    Benzeneacetonitrile, α-[3-[[2-(3,4-dimethoxyphenyl)ethyl]methyl
      amino]propyl]-3,4-dimethoxy-α-(1-methylethyl)-, labeled with
      deuterium (9CI)
MF    C27 H34 D4 N2 O4
```




```
HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):1
```

Structure matches are in line with the desired results.

Specify here the number of additional answers you want to see.

Run a Full File Search

- A full file structure search searches the entire database

```
=> S L1 FAM FULL

FULL SEARCH INITIATED 11:45:27
FULL SCREEN SEARCH COMPLETED -    265 TO ITERATE

100.0% PROCESSED     265 ITERATIONS              77 ANSWERS
SEARCH TIME: 00.00.01


L3            77 SEA FAM FUL L1
```

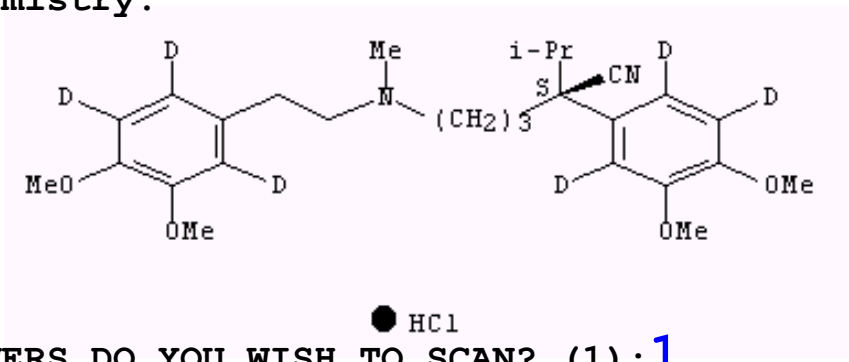
The results of a full file structure search are placed in a new answer set.

Run a Full File Search

- Option: Evaluate answers

```
=> D SCAN
L3   77 ANSWERS   REGISTRY  COPYRIGHT 2006 ACS on STN
IN    Benzene-2,3,6-d3-acetonitrile, α-[3-[[2-(4,5-dimethoxyphenyl)-2,3,6-
      d3-ethyl]methylamino]propyl]-4,5-dimethoxy-α-(1-methylethyl)-,
      monohydrochloride, (S)- (9CI)
MF    C27 H32 D6 N2 O4 . Cl H

Absolute stereochemistry.
```

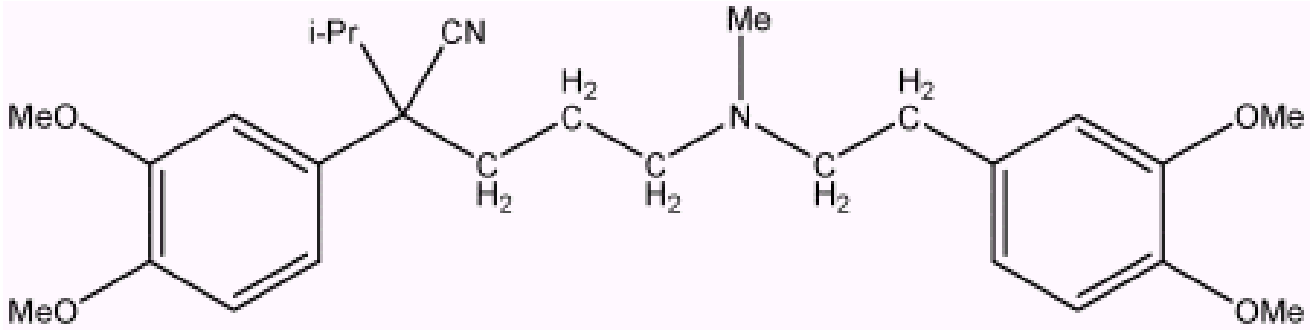


An isotopically labeled salt.

```
HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):1
```

Find Relevant References

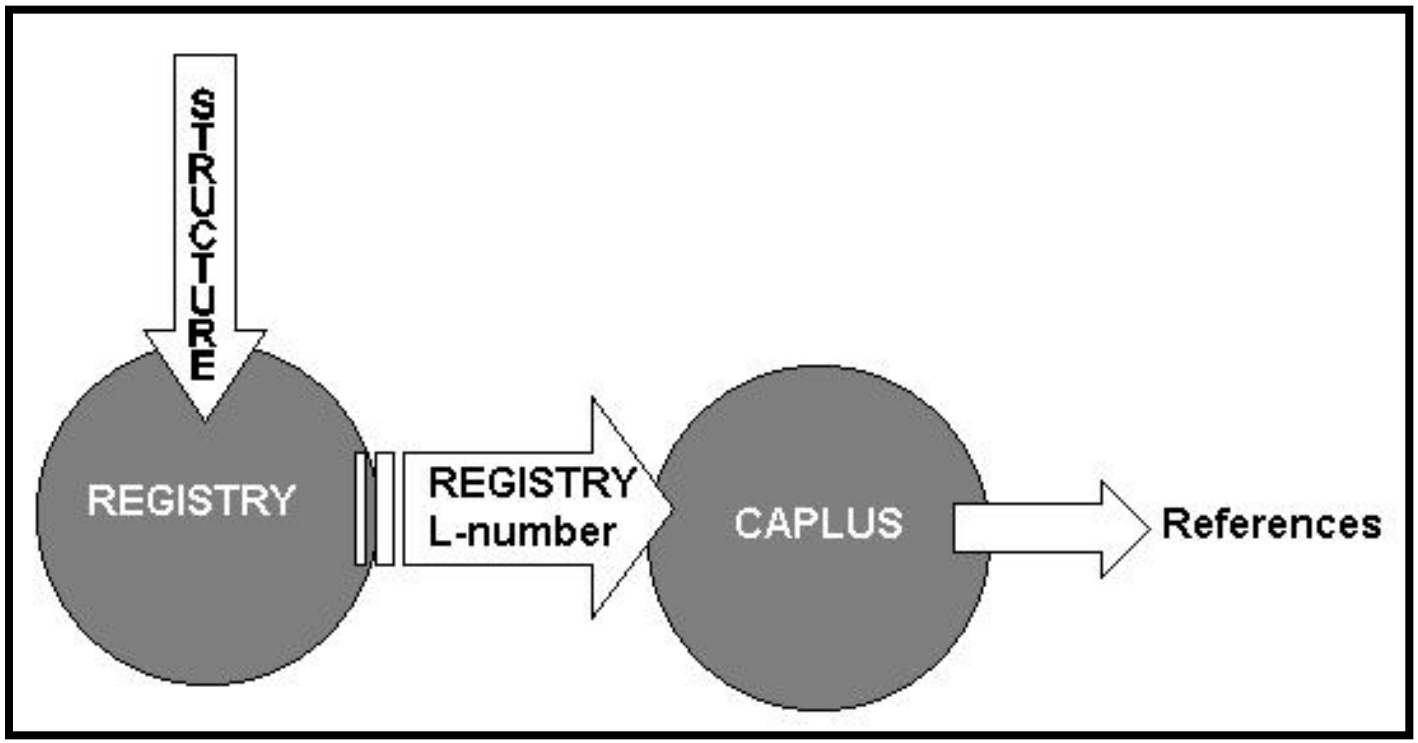
Recall the Search Question . . . *Locate references discussing the following substance, its salts, and any mixtures containing the substance:*



Find Relevant References

- REGISTRY contains chemical substance information
- Bibliographic references and abstracts of papers discussing substances retrieved by a structure search are stored in the CAplus database
- The L-number generated in the full file structure search is the key to locating relevant references

Find Relevant References



Find Relevant References

1. Enter CAplus

```
=> FILE CAPLUS
FILE 'CAPLUS' ENTERED AT 14:21:04 ON 23 MAR 2006
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2006 AMERICAN CHEMICAL SOCIETY (ACS)

Copyright of the articles to which records in this database refer is
held by the publishers listed in the PUBLISHER (PB) field (available
for records published or updated in Chemical Abstracts after December
26, 1996), unless otherwise indicated in the original publications.
The CA Lexicon is the copyrighted intellectual property of the
American Chemical Society and is provided to assist you in searching
databases on STN.  Any dissemination, distribution, copying, or storing
of this information, without the prior written consent of CAS, is
strictly prohibited.

FILE COVERS 1907 - 23 Mar 2006  VOL 144 ISS 13
FILE LAST UPDATED: 22 Mar 2006  (20060322/ED)
```

Find Relevant References

2. Search the REGISTRY L-number

```
=> S L3

L4          10112 L3
```


Find Relevant References

3. Evaluate results

```
=> D SCAN
L4    10112 ANSWERS   CAPLUS  COPYRIGHT 2006 ACS on STN
CC    4-3 (Toxicology)
TI    Expression and activity of a multixenobiotic resistance system in
      the Pacific oyster Crassostrea gigas
ST    multixenobiotic resistance system Pacific oyster; Crassostrea
      multixenobiotic resistance system
IT    Crassostrea gigas
      Development, nonmammalian postembryonic
      Environmental pollution
      Gill
         (expression and activity of multixenobiotic resistance system
         in Pacific oyster Crassostrea gigas)
IT    P-glycoproteins
      RL: BSU (Biological study, unclassified); BIOL (Biological study)
         (multixenobiotic resistance (MXR); expression and activity of
         multixenobiotic resistance system in Pacific oyster Crassostrea
         gigas)
```

(continued on next slide)

Find Relevant References

3. Evaluate results

```
IT   Egg
        (oocyte; expression and activity of multixenobiotic resistance
        system in Pacific oyster Crassostrea gigas)
IT   50-32-8, Benzo[a]pyrene, biological studies 52-53-9, Verapamil
     1912-24-9, Atrazine   59865-13-3, Cyclosporin
     RL: ADV (Adverse effect, including toxicity); BIOL (Biological
        study)
        (expression and activity of multixenobiotic resistance system
        in Pacific oyster Crassostrea gigas)

HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):1
```

This answer discusses the adverse effects of the compound.

Find Relevant References

3. Evaluate results

```
L4    10112 ANSWERS   CAPLUS  COPYRIGHT 2006 ACS on STN
IC    ICM  C07C255-43
      ICS  C07C253-34
CC    25-20(Benzene, Its Derivatives, and Condensed Benzenoid Compounds)
TI    Optical resolution of verapamil
ST    verapamil optical resoln
IT    56949-77-0
      RL: RCT (Reactant); RACT (Reactant or
         (optical resoln. of, method for)
IT    151079-03-7P 151079-04-8P
      RL: RCT (Reactant); SPN (Synthetic preparation); PREP
      (Preparation); RACT (Reactant or reagent)
         (prepn. and decompn. of, in prepn. of verapamil enantiomer)
IT    36622-28-3P, (S)-Verapamil hydrochloride 38176-02-2P,
      (R)-Verapamil hydrochloride
      RL: SPN (Synthetic preparation); PREP (Preparation)
         (prepn. of, method for)
HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):0
```

This answer discusses synthetic preparations.

STN is a unique partnership between two not-for-profit organizations to provide global access to the latest scientific discoveries

- Not-for-profit means that neither CAS nor FIZ Karlsruhe are driven by the monetary needs of shareholders
- Instead, both partners are wholly dedicated to serving the information needs of scientists

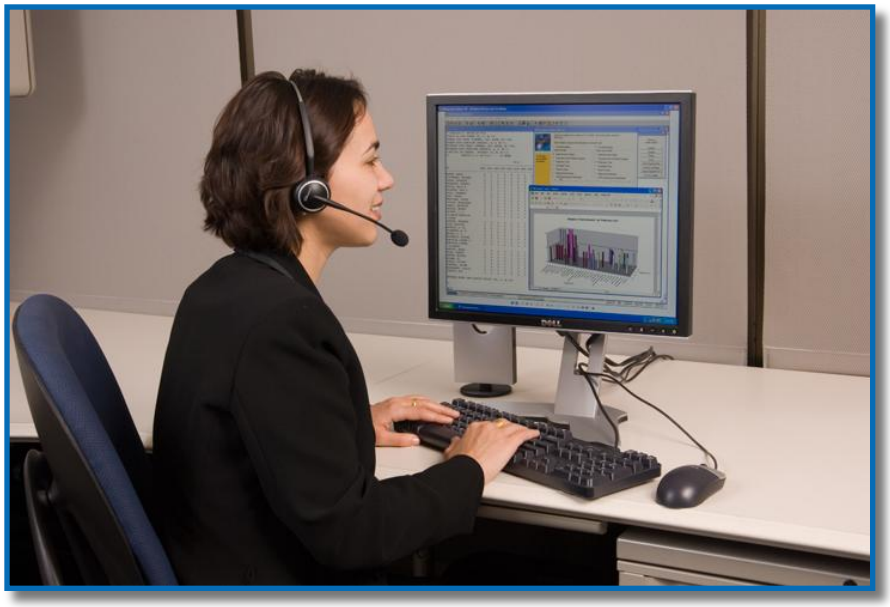
CAS

FIZ Karlsruhe

Scientists and highly skilled IT technicians staff our STN Help Desks in North America, Europe, and Asia

"You won't go wrong by contacting CAS folks. They are the most knowledgeable and most professional people that you can find out there. They will help you make the best decision for your needs."

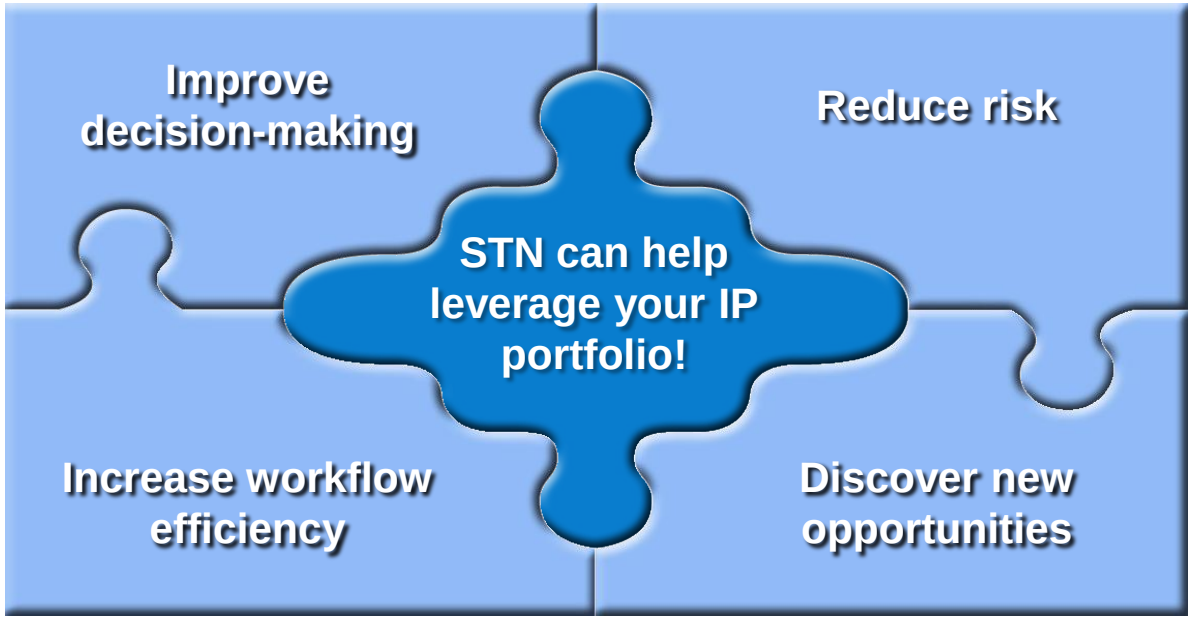
Rengin Konuk, Ph.D.
Knowledge Specialist
Ethicon, a J&J company
CHMINF list posting



STN has the right information for making business and project decisions

"STN offers all the important scientific databases we need. This allows comprehensive state-of-the-art searches on all our research topics."

Dr. Bernhard Hostettler, Head of Intellectual Property Management, Prionics, Swiss Animal Health Diagnostic Company



The Value of

for Business and Research Decisions

