

SciFinder Web - 2010

SciFinder File Descriptions - 2

Web System Recommendations and Requirements - 3

Chemical Abstracts Indexing Philosophy - 4

Tutorials, Search Examples, Search Tips, e-Seminars - 4

Comparison with other databases - 4

Web of Science

Reaxys: Beilstein/Gmelin/Organic Patents

Combined Chemical Dictionary

Medline

Explore References

Topic, Author, Company, Document - 5

Journal, Patent, Tags, Refining/Resorting/Displaying/Printing – 6

Explore Substances

Chemical Structure – 7

Markush - 7

Molecular Formula - 7

Substance Identifier (Registry Number, Name, Synonym) – 7-8

Refining/Resorting Substance Records – 8

Explore Reactions

Reaction Structure Searching – 9

Additional Information, Resources and Search Tips - 10-24

<http://library.caltech.edu/learning/classhandouts/SciFinderWeb2010.pdf>

SciFinder : a multi-file database:

Current database statistics: <http://www.cas.org/expertise/cascontent/ataglance/index.html>

SciFinder How to Guides: <http://www.cas.org/supduplicateport/scifi/htguides.html>

CAplus - The CAplus file(1a,17a) offers very broad subject coverage, extending to the chemical aspects of astronomy, biology, education, engineering, economics, geology, history, mathematics, medicine, and physics. It currently includes >33M records, adding 3K records per day.

CAplus coverage includes indexed abstracts of articles from journals (1907+) plus JACS (1879+), J. Phys. Chem. (1896+), J. Chem. Soc. Trans. (1878+), and landmark papers (1900-1912); ACS National Meeting Abstracts (1995+), JCS Abstracts (1890-1906) & Chem. Zentralblatt (1905-1906); web preprints (2000+); and regularly published conference proceedings, articles from one-time or first-time proceedings, dissertations, technical reports, technical disclosures, U.S. patents (1828+) and foreign patents (1907+) and patent family information (1957+) and edited research monograph chapters.

Bibliographic information and available abstracts are added for major country patents within 2 days, and articles from more than 1,500 key chemical journals within 8 days of receipt, prior to full CA indexing. The CAplus File is updated daily with about 3,000 new bibliographic records and weekly with indexing for about 14,000 records.

The CAplus file includes 200M cited references for journal articles, conference proceedings, and basic patents from the US, EPO, WIPO, and German patent offices,(1997+), and patent examiner citations for British and French patents (2003+).

CA Registry - The **CA Registry** file (3a) is a chemical structure and dictionary database listing >55M substances and >62M biosequences (adding about 12K/day). The REG file is adding both experimental [including links to spectra (17c)], and calculated physical property data on a continuous basis. (3b,17b)

Each substance is identified by a unique CAS Registry Number (RN). While many other databases and publications contain CAS RNs, they may be incorrect or out-of-date. The REG file is the final authority for CAS Registry Numbers. Each chemical structure may have multiple substances (isotopes, stereo variants, etc.) Exemplified prophetic substances (~160K) in selected patents have been recently added.

ChemCats - Links in REG to ~43M commercially available products (~12M chemicals) from over 1.1K suppliers & 1.2K catalogs. Supplier records provide a Catalog Name, Order #, Chemical Name, CAS RN, and may include Pricing Information).

CASReact - The CASReact file contains reaction information, for >40M organic and organometallic reactions (adding ~40K per week), from both the organic and organometallic sections of CA: journals(1985+), patents(1991+), InfoChem: journals(1974-99), patents(1982-99), INPI Core Reactions(1840-85), Biotransformations Database(1971-1998), EROS and the Wiley Reaction Collections (Organic Syntheses, Organic Reactions). Prophetic reactions have been recently added.

CHEMLIST - Links in REG to >275K chemical substances in national chemical inventories and regulatory lists. Each CHEMLIST record contains the CAS RN, a wide variety of international Chemical Names, File Segment, Confidentiality Status and Regulatory List Numbers, Inventory Status, U.S. EPA Regulations - TSCA, and summaries from the Chemical Regulation Reporter and the Federal Register, e.g., Report on Volume, Exposure, etc.), and a link 'Reaction Information' in CAS React.

MARPAT - Contains >800K searchable Markush (generic, hypothetical) structures from patents claims and disclosures (1961+), excluding: Alloys, metal oxides, inorganic salts, intermetallics, and polymers.

Medline - a journal article database, virtually identical to **PubMed**. Duplicate records can be removed (2a).

Web System Recommendations and Requirements

SciFinder General FAQ (includes links to Firefox FAQ, IE FAQ, Java FAQ, Structure Editor FAQ):
http://www.cas.org/help/scifinder/scifi_general_faq.htm

Windows-Based PCs:

CPU: 3GHz Core Duo
RAM: 2GB RAM
OS: Windows XP SP2
Browser: Firefox 3.5 or higher, IE7 or higher
Java Plug-in: Java 6 Update 16 or higher

Apple Mac OS X:

CPU: 2.66GHz Dual Core Xeon
RAM: 2GB RAM
OS: OS 10.6x
Browser: Firefox 3.5 or higher, Safari 4.x or higher
Java Plug-in: Latest JRE 1.6 bundled with MacOS 10.6

Internet Browser Requirements

JavaScript, Java, Cookies, Java Runtime Environment (JRE) update
Internet Explorer requires both:
 Appropriate ActiveX and plug-in permissions
 Responsive network connection (affects availability of structure drawing editor)
Add <https://scifinder.cas.org> to your browser "Trusted Sites" list.

Internet Connection Requirements - campus or VPN (21)

SSL (https) connection on port 443 to scifinder.cas.org
http connection on port 80 to chemport.cas.org for full text linking

PDF reader is required for printing.

Preview is Apple's free PDF viewer, built into OS X.
Adobe Reader can be used with Apple MAC or Windows based computers.
Adobe JavaScript must be enabled when using Adobe Reader or Adobe Acrobat Professional.

SciFinder System Availability

SciFinder is **off-line** from 7pm to 10 pm on Saturdays, except the 1st Saturday from 2pm – 10pm.

Saved Answer Sets, Combine Answer Sets, Keep Me Posted Results

Answer sets can be saved to the CAS Server (Save) or to the Desktop (Export).

Answer sets of the same kind: references, substances, or reactions can be combined after 'Saved'.

The 'Create Keep Me Posted' link (19) is available for either References or Substances searches.
<http://www.cas.org/support/scifi/etrain/kmp.html>

AutoSave feature saves your answer set if your session is inactive for more than 20 minutes, & are logged off.

CAS provides individualized help through help@cas.org or 1-800-753-4227.

Chemical Abstracts Indexing Philosophy

CA's policy of indexing the main points of an article can lead to unexpected results. For example, in an article on 'the effect of different cations on the IR spectra of Mo(CN)₈ complexes' (CAN 111:122893), each individual salt (e.g., Tripotassium octacyanomolybdate(3-) trihydrate) is indexed.

However, an article on the 'kinetics of the permanganate ion-potassium octacyanomolybdate(IV) reaction' (CAN 80:137539) only the Octacyanomolybdate(IV) ion is indexed, and not Potassium Octacyanomolybdate(IV).

In an article that references the IR spectra of 2,4-dimethyl-1,3,5-triazine' (CAN 85:93236), the synthesis of the compound is mentioned in the abstract, but the synthesis is not indexed, since the main point of the article was the IR spectra. In general, spectra entries are not generally assigned to articles where spectroscopic measurements are used merely for characterization (1b).

Tutorials, Search Examples, Search Tips, e-Seminars

Training material at <http://www.cas.org/support/scifi/index.html> is free to all. It includes:

SciFinder Tutorials – Web Version: References; Substances; Reactions; Additional Topics (recent enhancements, Overview, KMP profiles).

SciFinder 'How to Guides': Explore by; Working with Reference Answer Sets; Working with Substance Answer Sets; Working with Reaction Answer Sets; Additional Features (and newest features).

SciFinder Strategies: Small molecule searching; Synthetic chemistry; Polymer chemistry

SciFinder Solutions: Techniques and Tips <http://www.cas.org/support/scifi/sfsolutions/index.html>

SciFinder Talk (20): <http://listserv.cas.org/SCRIPTS/WA-CAS.EXE?A0=SCIFINDERTALK>

Comparison with other databases:

Web of Science is a journal article database (1900+) that also offers cited reference searching, and covers all areas of science, engineering and medicine. WoS added author abstracts in March 1992.

Reaxys consists of: **Beilstein** (chemical and physical properties, preparations, and reactions of organic compounds). It is especially useful for 1770-1980 literature since its coverage is currently focused on ~180 organic synthesis journals; **Gmelin** (chemical and physical properties, preparations and reactions of inorganic & organometallic compounds). While literature coverage is uneven (1c), it is useful for 1770-1975 references, since it is currently only indexing ~100 journals; and an organic **Patent Chemistry Database** (1976+).

Combined Chemical Dictionary includes over 540,000 compounds & their derivatives contained in the: Dictionary of Organic Compounds, Dictionary of Carbohydrates, Dictionary of Inorganic Compounds, Dictionary of Organometallic Compounds, Dictionary of Natural Products, Dictionary of Marine Natural Products, Dictionary of Analytical Reagents, Dictionary of Drugs, Dictionary of Food Compounds.

The CCD is very useful for basic physical property information, derivatives and unique literature references (e.g., Aldrich Spectra, preparation, spectra, hazard info, etc.).

SFS Medline (2b) is a journal article database, virtually identical to **PubMed**, that focuses on 'human health and disease' and indexes many of the chemistry, biochemistry, molecular & structural biology articles that are also indexed in SFS, and covers the literature since 1947. SFS Medline allows searching with Registry Number (1980+) as well as authors, keywords and MESH subject headings. Abstracts are available from 1976+. SFS Medline (in contrast with PubMed) does not provide links to ENTREZ sequence databases.

Explore References (Keep Me Posted available)

Research Topic Searching

Use keywords or a phrase that describes the topic (e.g. Treatment of autoimmune disorders). SFS has an internal dictionary which adds truncated & synonymous search terms (13a).

For example, phrases containing 'Cancer' are supplemented with 'neoplasms' and 'carcinoma' and phrases containing 'Freeze' are supplemented with 'froze', 'frozen', 'freezing', etc.

Search terms are automatically matched with names in REG File and CAS RNs are added as search terms (13b). CAS RNs can be used as keywords but this can be problematic (13c). Searching CAS RNs as substance identifiers (Explore Substances) is recommended.

CAS has a policy on searching code numbers, e.g., URB597 (11d). Research Topic searches can be pre-refined as shown.

Author Name Searching

Currently, author names appear exactly and as fully as presented in the original document. A misspelling will be corrected if it is recognized. Currently, all diacritical marks appearing in the original documents are eliminated (i.e., they are not transliterated). Practices, however, have varied over the years!!(14a).

Author names that do not appear in the Roman alphabet in the original source documents are transliterated as described in the "Introduction to the CA Author Index"(14b).

Variations in author names may occur in the database owing to differences in the original publications, e.g., initials for the first and/or middle names versus full first and/or middle names, or the use of different transliteration schemes (14c).

SFS has an internal dictionary which, by default, searches for alternative spellings of last name entered, including common and international variations. For example, "Smith" will retrieve: Schmidt, Schmitt, Smidt, Smit, Smith, and Smythe. This dictionary is obviously helpful for German names with umlauts (14d). For Harry Gray: H B, Harry, Harry B, Harry Barkus, Harry Barkus Jr.

'Company' Name Searching

Locate information associated with a specific company, university, government agency, or not-for-profit organization, which appear as the Corporate Source address (only for the first author). SFS has an algorithm for common organizational abbreviations. Both 'Caltech' and 'California Institute of Technology' retrieve the same number of references. Full company names (e.g., du Pont) are not phrase indexed.

Document Identifier Searching

Document identifiers include: CA/Medline Accession numbers (2009:624911/2003095979), CA document numbers (148:276656), Patent Numbers (JP 1992-502228), PubMed ID number (12608445), and Digital Object Identifier (10.1021/np050327j)

Journal / Patent Searching

Search for combinations of journal names (7b), either a full name, abbreviation, or acronym, and full or partial references, and/or title words and author names.

Search for Patent Numbers, Assignee Names (given in full but not phrase indexed), or Inventor Names (15b). Each member of a patent family (8a) is now searchable as a document identifier. Please note that patents are often given more descriptive titles (8b). Reading a patent record is described (8c).

Tags

Search for previously tagged records.

Refining/Resorting/Displaying/Printing

A list of references can be refined by checking-off individual records and then ... Keep Selected or Remove Selected. Remove Duplicates eliminates duplicate Medline records. 'Add Tags' allows personalized subject 'Tags' to be added to all or selected records.

Save (on the CAS server). Print offers a Print to PDF option. Export (save to your desktop). Copy/paste can be used for brief record printing.

Importing Data from SciFinder Scholar to **EndNote** is described at:
<http://instr.iastate.libguides.com/content.php?pid=5432&sid=34065>

If you Save your results to the CAS server you can NOT export that list of references in a tagged format.

The list can be resorted by author, publication date or title. Answers per Page can be reset from 15 to 100. The display can: hide, or display partial or full, abstracts. The Arrow inverts the order displayed.

Reference lists can be further refined by clicking either Analyze or Refine (13d): 'Analyze' to display a listing of Author Names, CAS RNs, CA Section Titles, Index Terms, etc. 'Refine' to limit by Research Topic (additional keyword terms), Document Type, Language, etc. There is also a 'Categorize' option (18b).

Either Select All or use the Check Boxes to 'Get Substances' [to retrieve Registry File records for substances in the article/s, which can be limited by 'Roles'], Get Reactions [to retrieve CASREACT records for the reactions in the article/s], Get Cited to display CA references cited in the article/s (1997+), Get Citing to display CA records that have cited the retrieved document/s since 1997 and Get Full Text.

Full-text article display depends on an institutional electronic subscription.

Each brief record also has options for: Substances, Reactions, Citing, Full Text, Comments, Tags

Clicking the **Title** of a reference displays the full CA record (bib info, abstract, citations (in articles published since 1997), CA indexing terms, including Patent Families (8a) and substances. The full CA record also includes cited patent examiner references from journal articles with text in the Latin alphabet and US, EPO, DE and WO patents, indexed from 1997+. Citations are linked to their CA record if available. CAS RNs are also linked to their Registry File records. 'Tags' and 'Comments' can be added.

Please minimize reading articles, as Caltech has only licensed 6 simultaneous users.

Explore Substances (Keep Me Posted available)

Chemical Structure Searching

Exact Structure Searching: <http://www.cas.org/support/scifi/etrain/exact.html>

Substructure Searching: <http://www.cas.org/support/scifi/etrain/sss.html>

Similarity Searching (9a): <http://www.cas.org/support/scifi/etrain/molsimilar.html>

Draw a structure, use a Template, convert an InChi or Smiles text string, or mouse-over a Substance Identifier structure in a Registry File structure (9a). Note: Atom, Shortcuts, Variables, and R-Group options, as well as the Repeating Group (26) and Variable Attachment(27) tools. Substitution and/or ring fusion/ring formation can be 'Locked Out' for most substances.

Clicking 'Show precision analysis' is strongly recommended (25). The 'Get Substances' radio buttons offer the choice of an 'Exact search' (no variable sites), a 'Substructure search' (some blocked sites may be overridden) or a Similarity search. Each of these search options also offer pre-limitations: Characteristics, Substance Classes, and Studies.

Markush Structure Searching

<http://www.cas.org/support/scifi/etrain/markush.html>

Explore Markush (24) patents by substructure and retrieve relevant patent references. 'Show full abstracts' to display Markush structures.

Molecular Formula Searching

Molecular formula searching in SciFinder Scholar is based on policies for the CA (print) Formula Index, which causes confusion (12a), e.g., CaSO₄ is searched as: < Ca. H₂ O₄ S >. More complex salts present similar problems (12b). Polymer searching is complicated by assignment of different Registry Numbers for essentially the same polymer depending on the polymerized monomer and how it was further processed (12c).

Substance Identifier: Registry Number and Chemical Name Searching

CAS Registry Numbers are simply accession numbers. K₃Mo(CN)₈ is a good example of the potentially confusing way 'substances' are registered(11a). In addition to 'plain' K₃Mo(CN)₈ (19442-23-0), CA has also registered K₃Mo(CN)₈.2 H₂O (123711-63-7), K₃Mo(CN)₈.3 H₂O (116188-33-1), and the anion Mo(CN)₈(-3) (17845-99-7). Searching 'octacyanomolybdate' will retrieve a wide variety of substances with different cations.

Similar problems are encountered with chiral compounds, since each isomer and the racemic mixture have different CAS RNs. (11b)

CAS RNs have been added to pre-1967 records but limiting with topic roles will eliminate these records.(11c)

Chemical Abstracts Index Names, other systematic chemical names, some name fragments, synonyms, GenBank Numbers, Enzyme Commission Numbers, Colour Index Numbers, etc. and trade names can be used as search terms.

For example, a search for Isopropanol labeled with O18 requires either the CA Index Name: 2-Propanol-18O or the trade name Isopropyl alcohol-18O. Some 'generic formula' substances are given CAS RNs: Propanol (generic formula) 62309-51-7 in addition to: 1-Propanol 71-23-8 & 2-Propanol 67-63-0.

Searching with an exact chemical name (e.g. Topotecan: 123948-87-8) does NOT retrieve Topotecan Hydrochloride: 119413-54-6. Searching for Nogitecan retrieves a record with the synonym Nogitecan Hydrochloride: 119413-54-6 since Nogitcan is not an exact chemical name. Note the inconsistent indexing of Codes, e.g. URB597. (11d)

Refining/Resorting Substance Records

A substance 'Answer Set' defaults to CAS RN order (which can be reversed by clicking the 'arrow') but can be resorted by: # of references, MW, or MF. Answers per Page can be reset from 15 to 50. The display can be viewed in 1,2,3 or 4 columns.

'Get References', for all or selected (use check-boxes) substances, can be limited by topic roles (but limits retrieval to 1967+, except for Preparation 1907+).

'Get Reactions' to retrieve CASREACT records for all or selected substances with options for limiting by reaction role (product, reactant, reagent, etc.).

'Get Commercial Sources' retrieves information from the ChemCats file.

'Combine Answer Sets' can be used if another answer set was previously saved. This is especially useful for 'NOTing' out unwanted structures. For example, conduct two substructure searches, the first for a complete answer set of reactions retrieving every possible option at the node of substitution, which is saved. In the second search, search for the reactants you don't want to have (e.g., halogens). Then use Combine to "not out" or exclude the second answer set from the first.

'Save' (saves the answer set on the CAS server),

'Print' (provides a PDF), and

'Export' (to your computer).

'Analysis' defaults to Commercial Availability, with additional options for: Elements, Reaction Availability, and Substance Role.

'Refine by' allows limiting the results by: modifying the Chemical Structure; include or exclude Isotopes; include or exclude: Metals; Commercial or Property Availability [i.e., any prop., any experimental prop., any calculated prop, or any specified (from a pick list of 10) experimental properties]; specific Property Values (9c) [e.g., boiling point/pressure ranges, melting point ranges, and ranges for all calculated (but not experimental) properties; Reference availability; and Atom Attachment [allows clicking on each atom for a list of attached elements/groups].

Each substance can be moused-over and click >> to: View Substance Detail, Explore by Structure, Markush, or Reactions; Synthesize; Get: Reactions (as a product, reactant, reagent, etc.); Commercial sources; Regulatory information; References), and Export as an Image.

Note links for 'Get References', 'Get Reactions', 'Get Commercial Sources', 'Get Regulatory Information', and a bookmark 'Link', 'Save', 'Print' and 'Export'.

Click on the 'Substance Detail/Registry Number' link to display the full Registry File record (CAS RN, MF, CA Index Name (the first listed name), Other Names, Class Identifier and Deleted CAS RNs (if appropriate), # of References, Document Types, CAS Roles (28), Predicted Properties and Experimental Properties (including links to NMR, IR, & Mass Spectra) with literature references (28).

CAS publishes the 'Ring Systems Handbook' in print that includes locant numbering (9b).

Explore Reactions

Reaction Structure Searching

Search for Reactions by Substructure: <http://www.cas.org/support/scifi/etrain/fullrss.html>

Explore Rxns by Specifying Rxn Sites: <http://www.cas.org/support/scifi/etrain/rxnsites.html>

Explore Rxns by Combining Answer Sets: <http://www.cas.org/support/scifi/etrain/combine.html>

Find Similar Reactions: <http://www.cas.org/support/scifi/etrain/similar.html>

Atom-Mapping, Solvent Limiters, and Functional Groups: <http://www.cas.org/support/scifi/tutorials.html>

CASREACT is a database of both single- and multi-step reactions. Click to open the structure editor, which offers additional 'reaction' tools: Lasso for duplicating structures ... drawing a reaction arrow ... adding reaction roles ... mapping atoms in a reaction ... mark bonds to be formed or broken ... searching for reactions of specific functional groups, as well as (keyboard) Shortcut Keys.

There are additional search options: Select Solvents (groups or individual); Select (non-participating functional) Group(s); Number of Steps; Classification/s (e.g. biotransformation, catalyzed, photochemical, etc.); Source/s (any, patents, non-patents); Publication Year/s.

There are two search options: 'Allow variability only as specified' (this refers to structures with variable atoms/groups and automatically locks out other substitution), or 'substructure' (this allows unlimited substitution in the drawn structure).

Reaction search results default to relevance (Tanimoto Similarity), but can be:

Re-sort by: Accession #; Experimental Procedure; # of steps; Product yield; Publication year.

Analyze by: Author Name, Catalyst, Company-Organization, Document Type, Experimental Procedure; Journal Name, Language, # of Steps, Product Yield, Publication Year or Solvent.

Refine by: changing the Reaction Structure ... minimum or maximum % Product Yield ... Number of Steps (single or range); by including or excluding 'Reaction Classification' (e.g. catalyzed, gas phase, photochemical, stereoselective, etc.); and all or any Non-participating Functional Groups;

There are options for 'Get References' (for selected reactions) and 'Find Additional Reactions'.

Each individual reaction record has links to: Reaction Detail, a bookmark Link and Similar Reactions (which has options for broad to narrow searching of transformation centers).

Each substance in the reaction diagram can be moused-over and click >> to: View Substance Detail, Explore by Structure, Markush, or Reactions; Synthesize; Get: Reactions (as a product, reactant, reagent, etc.); Commercial sources; Regulatory information; References).

The 'beaker' icon indicates commercial availability.

Additional Information, Resources and Search Tips

1a. CAplus – is a bibliographic file that augments Chemical Abstracts, in print (23), with additional records, such as: American Chemical Society National Meeting abstracts (5a) since 1995, 'abstracts' of JACS Communications 2002+, in-process records prior to full CAS indexing (6) for both articles (7) and patents (7), biographical material, book reviews, editorials, etc., from the [1,580 Core journals](#), that CAS has abstracted cover-to-cover, since October 1994, and over 224K [pre-1907 records](#) for journal articles, abstracts, and patents from 1808+. These records have been assigned volume “0,” OCI, and the year 1906, and include Abstracts and Section #s, but do not have CA indexing. Records from 1907+ include Abstracts, CA Index Terms (6), Registry Numbers and their original Section Numbers.

1b. Spectra Entries - CAS policy is NOT to make index entries for spectra when spectroscopic measurements are used merely as a means of characterizing new substances. There would normally have to be some novelty involved to index it. If you are looking for a literature reference, for example, to an NMR spectrum or chemical shift data for a particular compound and it's not in the major spectral compendia or indexed in the Scifinder REGfile (17c), it's possible that CAS has not indexed it ... as it's just "routine characterization". This policy extends to other spectra as well. Search options then would be 1) search for the earliest reported synthesis paper and check to see how the structure was identified, and 2) search for the compound in Beilstein and, for example, check for NMR field availability or 3) check the 'Finding Chemical Spectra and Spectral Data: Web Sources' web page at: <http://www.lib.utexas.edu/chem/info/spectra.html>

1c. The Gmelin Handbuch - Instead of providing comprehensive coverage of the literature over specific time periods, Gmelin offered comprehensive coverage of specific elements and their compounds. This approach resulted in very uneven coverage. Sodium covers the literature thru 1960-1971, but Potassium only thru 1936. The handbooks published after 1975 are not included, as Gmelin transitioned to uniformly indexing ~120 of the most important inorganic/organometallic journals beginning with 1976.

2a. Duplicate CAS/Medline records - While there is substantial duplication of records in SFS and PubMed, PubMed offers author abstracts which are generally more substantial than abstracts generated by CA staff. Automatic removal of duplicate Medline records is a 'Preferences' option. There is a 'Remove Duplicates' link on the References page and Refine by Database (CAPLUS or Medline) link.

2b. PubMed is an expanded online version of **Medline**, that includes out-of-scope citations (e.g. articles on femtochemistry, plate tectonics or astrophysics) from general science and chemistry journals (for which the life sciences articles are indexed for MEDLINE), citations that precede the date that a journal was selected for MEDLINE indexing and some additional life science journals that submit full text to PubMedCentral™ and receive a qualitative review by NLM. There is a subtle [difference between PubMed and Medline](#). SFS Medline includes **PreMedline** (in-process records without MESH indexing) but does not include [PubMed - as supplied by publisher] citations that are not indexed for Medline. Citations from [OLDMEDLINE](#) (1946-1965) are included in PubMed and SFS Medline.

3a. CAS Registry coverage is primarily substances identified from the scientific literature from 1957 to the present with some classes (fluorine- and silicon-containing compounds) going back to the early 1900s. Records from 1907-1967 contain CAS Registry Numbers for identified compounds, but do NOT contain all CAS Roles. Restricting a substance search (with the exception of PREP) will limit retrieval to records from 1967+.

A [CAS website](#) gives the most recent CAS registry number, the latest count in chemical substance registrations in REGISTRY, reactions in CASREACT, commercially available compounds in CHEMCATS, and inventoried/regulated substances in CHEMLIST.

3b. Experimental & Calculated Property Groups (22)

ACD calculated property data is provided for over 15M organic compounds, and over 1M experimental physical properties (e.g. MP, BP, ORP, RI) were added from the SPRESI database (1975-1991) for over 830K compounds.

Experimental property data from CAS abstracts is also being added. Beginning in late March 2005, over 2 million substances in REGISTRY will be further enriched with nearly 5 million references or "tags" that point to additional experimental property data. The tags for 186 different properties will have bibliographic references that direct you to the original source documents. There are ~2B predicted/experimental data tags.

The tagged properties can be spectra, charts, graphs, large tables, or numeric values. The tags point to the literature reference where those data may be found. Tags will continue to be added as new references containing property data are identified.

Links to actual spectra are provided for ~150,000 compounds. IR, MS and ¹³C NMR spectra are from Wiley spectral databases and the Japanese SDSB database, in addition to ~24M predicted ¹H NMR Spectra.

In order to provide the best reliability in the **ACD calculated values** (17b), CAS chose limits for some of the values. For example, there is a limit of 0-14 for pKa values. Since 2-phenylethanol (for example) has a pKa of about 15, its pKa is not listed at this time. A full **description of the sources** for experimental and calculated properties is provided along with a **properties glossary**.

4a. CHEMLIST (Regulated Chemicals Listing) was originally built from data in the 1985 TSCA Inventory and supplements, and the Federal Register for 1978-1987. Sources for the weekly updates have been greatly expanded (e.g. EINECS, EPA, etc.). CHEMLIST contains records from the national chemical regulatory inventories of the United States, Canada, Europe, Australia, Japan, Korea, and the Philippines(4a), from 1979+. It is also a good source for foreign chemical names/synonyms, e.g. Tetrachloroetileno(Spain) = Perchloroethylene(U.S.)

4b. CASREACT has recently added 99,000 reactions from approximately 54,000 documents published between 1907-1985, from the **INPI Core Reaction Database**, which is rich in preparations of steroids, heterocyclic compounds, and carbohydrates. This database contains proven synthetic methods compiled by research scientists at the French pharmaceutical company Roussel-Uclaf and covers reactions from both scientific journals and patents.

5. Abstracts of Papers ... ACS National Meetings - CAS began adding records for the ACS National Meeting abstracts to the online CPlus file (SciFinder Scholar) beginning with the 210th National Meeting, August 1995. These are CPlus records only. The print version of CA does not include records for the abstracts of from the ACS national meetings (nor any document that is 'abstracts only').

Please note that Web of Science has been indexing Abstracts ... ACS Meetings since 1969.

CPlus also currently indexes National Meeting preprints from the following five ACS technical divisions:

Division of Environmental Chemistry

Division of Polymer Chemistry

Division of Polymeric Materials: Science and Engineering (PMSE)

Division of Fuel Chemistry

Division of Petroleum Chemistry

J. Chem. Soc. Abstracts (1890-1906) provide an extensive English language abstract. Chem. Zentralblatt (1905-1906) titles and abstracts are given as machine translations.

6. A CA Index Guide is published for each Collective Index period beginning with the 8th (1967-71). Each Index Guide is composed of a main body and four appendixes. The main body contains cross-references, indexing notes, valid general subject index headings, and diagrams for cyclic and acyclic stereoparents (that were previously included in the CA Subject Index (1907-1966)). Appendixes 1-3 deal with index headings and policies. Appendix 4 is a definitive listing of rules for assigning Chemical Abstract Index Names.

7a. Indexing CAS Core Journals & Patents - For the ~1500 "core" chemical journals (e.g. JACS, Chem. Commun., etc.) < <http://info.cas.org/sent.html> > bibliographic and abstract information is generally available in CAPLUS (SciFinder Scholar) within 8 days of receipt by CAS. Indexing is progressively added to the record. This explains why an author search will retrieve some very recent papers, while a substance search will not. These 'incomplete' records are highlighted with an "INDEXING IN PROGRESS" note in the record. Indexing for these journals is generally complete in 30-90 days. For patents from the primary patent offices (US, EP, WO, JP, UK, FR and DE) bibliographic and abstract (machine translations for JP and DE) information is available within two days of publication by those patent offices and full indexing is completed within 30 days.

7b. The Chemical Abstracts Source Index (CASSI) is the complete list of publications indexed by CAS (in print only) and is supplemented by the **CAS DDS file**, which is a searchable (title keywords, CODEN, ISSN) list of publications currently indexed for SFS.

There is an online list of the 1580 core journals abstracted by Chemical Abstracts at: <http://www.cas.org/expertise/cascontent/caplus/corejournals.html>

8a. Patent Family information is a convenient way to find U.S. equivalents, since CA generally has abstracted European, German or WO applications first. Esp@cenet also provides patent family information

8b. Patent Title Changes - CAS editorial staff feel that, in many cases, the original title of many patents are not explicit enough to be useful to searchers. CAS may therefore revise the title to provide more descriptive information. For example:

Dervan, et. al. Dec. 7, 1999. US 5,998,140

Complex formation between dsDNA and oligomer of cyclic heterocycles

Dervan, Peter B. US 5998140. Chemical Abstracts 132:19609

Inhibition of transcription or cell proliferation with DNA-binding polyamides.

8c. How to Read Patent Information in SciFinder Scholar

9a. Copy/paste a structure – Locate a substance (e.g. strychnine) in the REG file, and mouse-over on the structure diagram, and then click the >> to 'explore by chemical structure', etc. The structure can be modified and then searched. **Similarity search** uses a Tanimoto algorithm.

9b. The **Ring Systems Handbook** (MIL-8 QREF) lists the CA index name, CAS Registry Number, MF and structural diagrams (with locant numbering) for all ring systems indexed in the CAS Registry File. Search by CA Index Name, ring analysis or MF. Using SciFinder, you must build a parent ring structure (locking out substitution) and run an Exact Structure search. SciFinder lists other names (synonyms), but omits locant numbering. The Handbook conveniently groups all similar ring structures together so that you can quickly scan structural diagrams for additional compounds of interest.

9c. **Limiting with 'Property Data' includes:** Experimental BP/Pressure and MP, Calculated # of hydrogen donors or acceptors, MW, Log POW, Freely Rotatable Bonds, Bioconcentration Factor, BP, Enthalpy of Vaporization, Flash Point, Koc (as a function of pH) Log D (a function of pH), Molar solubility (as a function of pH), pKa and Vapor Pressure.

Limiting with 'Property Availability' includes: any property, any calculated property, any experimental property or any selected experimental property. See (3b) for additional information on properties.

10. Retrieving References /Cited References and Citing Reference Searching

Cited references are included in the CAPLUS database from all articles (in journals published in the Latin Alphabet) from the nearly 1300 core journals; indexed articles from 5300 other journals and miscellaneous publications; and basic patents (examiner citations only) from the U.S., Germany, EPO, and WIPO since 1997.

A unique capability of the CAS citations is the linking from the cited references to the corresponding CAplus or Medline abstract, which provides more complete information about the research in the citing document. The typical cited reference includes the first author's name, title of the referenced journal, patent, or other publication, year of publication, volume number, and beginning page number. CAS lists each cited reference, even if duplicated, while WoS only lists duplicated cited references once and does not list databases, etc. SciFinder has a correction mechanism for commonly miss-cited journal titles (e.g. JCS Chem. Commun. ⇔ Chem. Commun.). Cited references from non-Roman language journals, (e.g. Yuki Gosei Kagaku Kyokaiishi = J. Syn. Org. Chem. Jpn.) are not included.

11. CAS Registry Number / Chemical Name Searching

11a. Searching for Octacyanomolybdate compounds with a CAS RN (e.g. 19442-23-0 for $K_3Mo(CN)_8$), does not easily lead to the compounds with other cations (e.g. Na^+ , Li^+ , Rb^+ , Fe^{++}). All of these other compounds could be missed because there is no inter-linking of CAS Registry Numbers. This is a general 'problem' in SciFinder Scholar as derivatives of organic compounds are also NOT linked to their parent compound as they are in the [Combined Chemical Dictionary](#). Problems are also encountered with hydrates & minerals/synthetics.

One way to retrieve all possible compounds of interest is to first search using the molecular formula $C_8MoN_8.3K$. The registry record for this compound, provides the CA Index Name. Copy the main part of this name: Molybdate(3-), octakis(cyano-C)-, and search this 'name' as a 'substance identifier'. SciFinder is programmed to treat a string of characters in this field as a truncated search term, and retrieves a variety of salts and the free ion (which has a unique CAS RN).

There is a further complication when searching for either salts or coordination compounds of organic acids. For example, searching for analytical techniques for the determination of Aurintricarboxylic Acid or its salts and coordination compounds, with only the CAS RN 569-58-4 for the tri-ammonium salt, suggests a substance identifier search for either 'Aurintricarboxylic' (retrieves three compounds with references, Barium and ammonium salts, and the free acid) or 'Aurintricarboxylate' (retrieves 8 compounds (5 rare earth coordination complexes (e.g. 75010-07-0) and the Beryllium, ethylenediamine and ammonium salts). Since the CA CSI (in print), lists all salts and coordination compounds together, one now finds the Sodium salt., which didn't show up in the previous searches because the only synonyms in the REG file are: C.I. Mordant Violet 39, trisodium salt (8CI); C.I. 43810; C.I. Mordant Violet 39; Chrome Violet G; Mordant Violet 39.

Copying the structure of the Ammonium salt into the 'structure editor' and deleting the 'ammonium ion' and, doing an exact search, only retrieves the ATA Polymer and the Acid. Including 'component of a larger system', retrieves the substances mentioned above (plus some more) but not the coordination compounds. Searching with the structure copied from a coordination compound (e.g., 75010-07-0) and changing the metal to 'any metal' retrieves 28 substances (with many having synonyms named as complexes (e.g. Benzoic acid, 5-[(3-carboxy-4-hydroxy-5-methylphenyl)(3-carboxy-4-oxo-2,5-cyclohexadien-1-ylidene)methyl]-2-hydroxy-3-methyl-, **holmium complex**) instead of: Gadolinium **aurintricarboxylate**, the form used for the 5 retrieved above. This again is a reminder of the utility of the 'print' CA Chemical Substance Index.

11b. Searching may be further complicated by difficulties in easily retrieving CAS RNs for (+) and (-) isomers. A quick search of the CA Chemical Substance Index (in print), reveals the registry numbers for: (-)-Isoborneol, (+) Isoborneol, which can also be found in SFS, by searching with this form of the name. In general, finding all isomers may require a structure search. Each stereoisomer is a 'different' compound with a different CAS RN. For compounds with multiple stereo-centers, CAS designates the racemic mixture with (-rel) at the end of the stereo term, e.g., (2R,3S)-rel- which represents both the (2R,3S) and the (2S,3R) isomers. This is in contrast with the CCD which would have used (2RS,3SR).

Please also note that the default retrieval is generally the racemic (+/-) mixture. Beginning in 1997, racemic mixtures of substances having single chiral centers are now indexed, registered, and named as [non-stereospecific substances](#). The decision to discontinue the racemate distinction was based on the observation

that most of the chemical substances encountered in the literature are presented without any explicit indications of their racemic nature but which in the majority of the cases are probably racemates. The distinction that CAS was making between the racemic and non-stereospecific (or relative) forms was judged to cause scattering in indexing and registration, and complications in database searching.

Thus, DL-malic acid and malic acid (stereospecificity unspecified) now receive the same CAS Registry Number and CA index name: 6915-15-7, Butanedioic acid, hydroxy-. The Registry Number for the DL-form (617-48-1) will be cross-referred to the Registry Number of the non-stereospecific form (6915-15-7).

Racemates having more than one chiral center are indexed, registered, and named as having only relative stereochemistry. Thus, DL-threitol and threitol (absolute stereospecificity unspecified but having two chiral centers with the same relative configuration) now receive the same CAS Registry Number 7493-90-5.

11c. Pre-1967 records in SciFinder Scholar contain CAS Registry Numbers for identified compounds. Unidentified compounds can only be searched with names in 'Research Topic'. Searching for DDT (50-293) and retrieving 'references associated with' roles such as 'Analytical Study', for example, automatically limits retrieval to references indexed from 1967+. Searching for 'analysis of DDT' as a 'Research Topic' (i.e., mentioned in the title, abstract or indexing phrase) retrieves all references. For exhaustive searching (1907-1966), the (print) CA indexes/abstracts should be used for MF, compound name and subject searching.

Please be careful searching with CAS Registry Numbers as Research Topic subject terms. Searching for 124-76-5p (the CAS RN for Isoborneol, limited to preparation references) retrieves 149 references. Searching Isoborneol as a substance identifier (Explore by Chemical Substance) and limiting to 'preparation' retrieves 157 references. The difference is due to a previously used but deleted registry number that is only retrieved thru a 'substance identifier' search (since this will search for both numbers).

Further problems are encountered when searching with CAS RNs for substances that are currently indexed as concepts (i.e., General Subject Index headings). For example, Mica was indexed in the CA CSI (with CAS RN 12001-26-2) thru 1981. Since 1982, it is indexed in the CA GSI under 'Mica Group Minerals'. The REG file record for Mica gives the CAS RN, but the CA Index Name is given as 'Mica Group Minerals'. Clicking on 'Get References' for 12001-26-2 retrieves 733 records (which come from flagged RNs in the abstracts thru 1981). The CAS RNs were removed from the index terms in the pre-1982 records so that to search for 'Mica' records prior to 1981, you have to search it as a Research Topic.

11d. CAS has a standard policy on Code #s. 1) Letters followed by numbers require a space between [URB597 --> URB 597]. 2) Punctuation between like [numbers-numbers or letters-letters] is retained, although 1,000's commas are removed. 3) Numbers followed by letters are required to be closed up as are all subsequent characters except as noted in 2. 4) Punctuation between unlike [letters-numbers or numbers-letters] are removed except as in 1 above. Where the numbers are clearly specified locants, hyphens are retained; e.g., 2,4-D is the best example.

12. Molecular Formula Searching

12a. **Simple salts** like sodium chloride are searched as: < NaCl > (or more properly alphabetically) as: < ClNa > More complex inorganic salts, like Calcium Sulfate or Phosphate, must be searched as: < Ca.H2 O4 S > or < Ca.2/3 H3 O4 P >. This reflects the print volume policy, where ALL the salts of Sulfuric or Phosphoric Acid, for example, are listed together under H2O4S or H3O4P. This policy also applies to simple organic salts like Sodium Benzoate which, however, are searched as: < C7H6O2.Na >, but again the MF for the acid is searched.

12b. **Complex organic/organometallic salts** are searched by entering the (MF of the cation).x(MF of the anion), where x=the number of anions. For example, Tris(2,2'-bipyridine)iron(2+)bis(tetrafluoroborate) is searched as: < C30H24FeN6.2BF4 > and Tetrakis(tetrabutylammonium)octacyanomolybdate is searched as: < C16H36N.1/4C8MoN8 >

12c. **Polymer molecular formula searching:** Simple homopolymers (e.g. Polystyrene) are searched as: $(C_8H_8)_x$. The results may need to be refined by searching for the 'phenyl ring' as a substructure. For a more complex polymer, e.g. with internal repeating groups: $HOOCCH_2CH_2O-[CH_2CH_2O]_n-CH_2CH_2COOH$, is searched as $(C_2H_4O)_n C_6H_{10}O_5$. This points out an additional problem, namely that a CAS RN is assigned not only for the polymer but also for each compound where n is known (e.g. n=1, n=2, etc.) and for each stereoisomer. There is a POLYLINK command which

Simple homopolymers (e.g. Polystyrene) are searched as: $(C_8H_8)_x$. Results may need to be refined (e.g. searching for the 'phenyl ring' as a substructure and eliminating isotopic variations). For a more complex polymer, e.g. with internal repeating groups: $HOOCCH_2CH_2O-[CH_2CH_2O]_n-CH_2CH_2COOH$, is searched as $(C_2H_4O)_n C_6H_{10}O_5$. These examples point out an additional problem, namely that a CAS RN is assigned for the polymer and also for each compound where n is known (e.g. n=1, n=2, etc.) and for each stereoisomer (syndiotactic, isotactic, etc.). There is a [STN POLYLINK](#) command which retrieves related polymers

Polymer searching is further complicated by assignment of different Registry Numbers for essentially the same polymer depending on the polymerized monomer and how it was further processed. For example, the polymer from the sodium salt monomer of para-vinylbenzenesulfonic acid (25704-18-1 has the MF= $(C_8H_8O_3S.Na)_x$, indicating 100% salt formation. The polymer $(C_8H_8O_3S).xNa$, indicates the polymerization of the sulfonic acid monomer, and post-treatment to create the salt (28038-50-8). Depending on the conditions of the post-treatment, the degree to which the pendant sulfonic acid groups would be converted to sodium salts would vary, and CAS does not assume complete salt formation. This results in an "unspecified" salt (e.g., x Na, where x represents an unknown number) since the exact nature of the substance is not known.

Specific terms, such as diblock, triblock, tetrablock or pentablock, are now used for all block polymers registered from 1987 to date. Stereoblock was added, when appropriate, for block polymers registered from 1967 to the present.

13. **Research Topic Searching** (truncation symbols are NOT recognized)

SciFinder Scholar 2006: An Empirical Analysis of Research Topic Querying JCIM 2006, 46(2), 767-774

13a. Single words and phrases are searched as entered, and results also allow selecting reference sets also containing CAS' synonyms (i.e. containing the concept). Phrases should be entered as complete sentences in English. SciFinder Scholar understands which terms are the key words and the conjunctions/prepositions determine how these terms relate to each other. Please note that prepositions are place holders. Searching for: 'thin films in semiconductors' searches for the exact phrase and a Boolean search for 'thin films' and 'semiconductors' both closely associated and anywhere in the record. Research topics can be composed of two or three concepts combined with prepositions, conjunctions, etc. For example:

I am interested in liver apoptosis and nasal apoptosis in cancer.

Prepositions

Examples of common prepositions are: after, among, at, between, from, in, into, on, upon In English, there is a difference between the two phrases:

Reactions caused by heat --- Reactions causing heat

Since SciFinder Scholar does not understand the relationship implied by caused by or causing, it will display results for the exact phrases, along with identical results containing the concepts 'Reactions' and 'Heat'.

Use of OR and AND

Sometimes, SFS cannot differentiate AND and OR. In these cases, SciFinder Scholar will run the search and present both results. Do not use the words AND and OR if a preposition would be more exact.

Instead of: "I am interested in human growth hormone and fetal development"

Use: "I am interested in the effects of human growth hormone on fetal development"

Do not use the word AND when you really mean OR.

I am interested in interface software for bibliographic information AND numeric data
(This means that you want a single piece of software to interface both with bibliographic information and with numeric data.)

I am interested in interface software for bibliographic information OR numeric data
(This means that you are interested in interface software for either alternative.)

Distributed Modifiers

SciFinder Scholar does not distribute modifiers. "I am interested in liver or nasal apoptosis", must be expanded to: "I am interested in liver apoptosis or nasal apoptosis".

Negation

You may describe a research topic using negative terms such as "not" or "except". Example:
I am interested in ringed planets but NOT Saturn

Abbreviations and Misspellings

SciFinder Scholar is designed to recognize and accept:

Commonly used abbreviations, such as "BTU" and "Prep"
Commonly misspelled words ("affect" instead of "effect")
Words spelled according to British English ("colour" instead of "color")

Synonyms

Place synonyms (limit of 3) in parenthesis next to the topic you entered. Example:

I am interested in the milk production of cows (bovines)

General Subject Headings

If a query exactly matches a CA General Subject Heading, e.g. 'Carbon Fibers', no additional processing is done. Searching 'Carbon Fiber' would give additional results.

Truncated Words and Truncating Symbols

SciFinder Scholar automatically truncates terms and searches for plural forms, past tense verbs, etc., unless searched as a single term. Example:

What are the effects of age on calcium absorption in women?

SciFinder Scholar automatically expands the query to include terms such as 'aged' and 'aging.' Singular forms of plurals are NOT searched. SciFinder will disregard truncation or wildcard symbols and interpret the remaining characters literally.

Boolean Searches

Do not use Boolean logic or wording groupings, as parentheses indicate synonyms (with a limit of three), not groupings. Use conjunctions and prepositions instead. An exception is (NOTL) ... NOT proximity operator.

13b. SciFinder Scholar applies many different **algorithms and dictionaries** in the background when searching. In the case of the Research Topic Explore, the algorithm actually tries to match *everything* you input in your phrase (specifically, the longest name strings possible) against Registry for a name match. For whatever match or matches it finds, it will then include the RN (plus any deleted, alternate, etc. RNs) from that record along with the actual term as part of your Research Topic search.

13c. In a Research Topic search for articles on the reaction of 'dicobalt octacarbonyl' and 'tributylphosphine', both compounds have a **name match in Registry**. The search performed that retrieved 143 references was actually:

(dicobalt octacarbonyl or 10210-68-1) and (tributylphosphine or 998-40-3)

Including RNs in the search statement is not recommended. This, because the algorithm can get "confused" by the variations in logic. The following search, retrieved only 121 references, since all terms are and'ed:

Dicobalt octacarbonyl(10210-68-1) and Tributylphosphine(998-40-3)

Use of a preposition (such as with) to "link" concepts -- even if that may not look grammatically correct, often offers additional combinations. For example:

dicobalt octacarbonyl with tributylphosphine

not only searches both the names and RNs but also displays additional result sets:

- 6 references were found containing "dicobalt octacarbonyl with tributylphosphine" as entered. Please note that the 'as entered' search does NOT include registry matches.

- 45 references were found containing the two concepts "dicobalt octacarbonyl" and "tributylphosphine" closely associated with one another.

13d. **Analyze by:** Author Name, CAS RN, CA Section Title, Company/Organization, Database, Document Type, Index Term, Journal Name, Language, Publication Year, Supplementary Term ... All or selected documents, and sort alphabetically or by frequency. See (18a) for a detailed analysis.

Refine to limit by:

Topic, Company Name, Author Name, Publication Year, Document Type, Language, Database.

Get Related:

Cited References, Citing References, Substances Indexed (that can be limited with topic 'Roles', which limits search to 1967+), Reactions Indexed, Information from the WWW.

13e. **Stop words** are included in phrase searches if not separated by prepositions. For example;

Quantum Size Effect 3,442 as entered, 4,814 as concept

Effect of Quantum Size 180 as entered, 10,573 as concept (since 'effect' is a stop word)

14. Indexing and Editing of Author Names:

14a. Historically, Chemical Abstracts presented author names (in Roman script) exactly as they appeared in the original documents complete with diacritical marks. The practice of transliterating letters with diacritical marks and special characters with English equivalents was introduced in 1961. Thus, for example, the Germanic letter "u" with an umlaut was replaced by "ue". In 1997, transliteration of characters with diacritical marks was discontinued and the diacritical marks were simply removed from the proper names. This change in practice occurred because CAS' ability to provide consistent transliteration of author names had been declining as more and more material was received in electronic form from publishers with varying treatments of such characters and because of different editorial practices in the original literature.

14b. Chinese Names: A change in transliteration of author names occurred in 1982 when CAS began to use the Pinyin system for transliterating author names appearing in publications originating in the People's Republic of China, while retaining the Wade-Giles system for Chinese-language documents from Taiwan and other countries. This change was made because the Pinyin scheme had been adopted by the PRC as its official method of transliteration. The use of these two Chinese transliteration schemes continues today.

14c. The transliteration of Cyrillic names, to the Latin alphabet, is complicated by the use of three different schemes (in the U.S. alone) and further problems arise in German and French indexes (1a). Zhanov, for example, in an English language index may be written as Shdanow in a German language index and Jdanoff in a French language index. A comparative table of transliteration schemes is available (1b). There are also Japanese transliteration variations (e.g. Oe or Ohe) that depend on personal preference. SciFinder Scholar has a 'alternative spelling' algorithm that retrieves records for Lyubinskii, Lyubinsky, Liubinskii or Liubinsky when searching for any one of the four possibilities.

PubMed	CA(print)	WoS	SFS
Izvestiia	Izvestiya	Izvestiya	Izvestiya
Liubinskii	Lyubinskii	Lyubinsky	L(i/y)ubinsk(ii/y)

14d. Author Indexes (in Print): Throughout much of its early history, CAS editors sought to expand initials in author names when compiling the printed volume and collective author indexes to provide the full author names and thereby avoid scattering of entries for papers and patents for the same author or inventor. This labor intensive practice, which affected only the printed indexes, was discontinued in 1992. Prior to 1997, CAS only indexed the first 10 authors (9 and et. al., if more than 10). Beginning in 1997 all authors are listed. The practice of listing only the address of the first author continues. As an aside, Ingenta may only index the first, second and last authors, while Web of Science indexes all authors & lists all author addresses. Only the first author's address in WoS is directly linked to a name. Other addresses are listed in the order they appear in the article.

Please note that the filing order in the print CA Author Indexes is by initials (not by full forenames).

15. Document Identifier Searching

15a. Prior to 1967, CAS used volume/page or volume/column numbers with letter designations to pin-point abstract location in the column. These 'abstract' numbers are found in literature references and handbooks, and are currently NOT searchable in SFS. The print volumes of CA from 1907-1966 have been retained on Mil-8.

15b. All patent numbers and patent assignees from 1907+ and patent family data from 1957+ are easily searchable. Nearly 800,000 abstracted patents and over 83,000 family members are included for the years 1907-1966.. For example, SFS lists 13,157 patents assigned to Merck between 1907-1966.

16. Company Name / Organization Searching

Search for Merck as a Company Name, Refine references as Patents, and either Analyze by publication year for a histogram of patents by year or Refine again by year or range (e.g. 1907-1941). Algorithms allow searching for Caltech or California Institute of Technology.

17. Enhancements

17a. Subject and substance indexing has been added to more than 11,500 pre-1907 records in CA/CAPLUS. Records included in this enhancement represent original articles, and article and patent abstracts from the following journals: JACS, JPC, JCS Trans., JCS Absts. Many of the U.S. patent abstract records have been replaced with patent records that include more extensive bibliographic and abstract information. CAPLUS has been enhanced with approximately 13,000 traditional medicine patents from 1985+. These patents are not represented by new or different kind codes and are not assigned to a CA section.

17b. The Rule-of-5 (Ro5) has been widely adopted in the drug discovery industry (has similar equivalents in the agrochemical world). This rule of thumb is used as a first step in drug discovery to quickly eliminate lead candidates with poor bioavailability properties. One of the key parameters of the Ro5 is logP—a useful descriptor but one that fails to take into account the impact on lipophilicity of varying concentrations of ionic species in biological systems. ACD/Labs has compared application of the Ro5 using logD in place of logP on 9 commercial compound libraries.

The Rule of Five Revisited. *Molecular Pharmaceutics*; 2007; 4(4); 556-560. DOI: [10.1021/mp0700209](https://doi.org/10.1021/mp0700209)
Christopher Lipinski C&EN November 5, 2007 p.18

17c. Experimental Spectra: Since the beginning of 2008, over 700K experimental spectra have been added to REGISTRY, including: 408K proton, carbon-13, and heteroatom NMR, 127K mass spectra, 174K IR and 3K Raman.

This enhancement is part of an ongoing effort to strengthen the collection of experimental spectra first made available in 2005. REGISTRY now includes more than 480,000 experimental spectra.

18a. Analysis:

CA Sections, Index Terms, Concept Headings, Supplementary Terms, Registry Number, Concepts

<http://www.cas.org/products/print/ca/casections.html>

1) CA Section Title has its origins in the original print Chemical Abstracts which appeared in 80 sections to help people browse. These are very broad categories. Note that the definitions and exact title of the sections has changed a number of times, over the years, which explains the variations you will see when you do an Analysis. Note that doing this analysis automatically discards Medline records (with no warning message).

2) Index Term - analyzes the controlled vocabulary of both CAPLUS and MEDLINE, i.e. subject headings, but not the chemical substance indexing. It does NOT search the Supplementary Terms.

3) CA Concept Headings - this analyzes CA "main heading" controlled vocabulary/index terms used in the print General Subject Index (which excludes chemical substance indexing). These headings appear in the CONCEPT column (the header box, not the detailed text modifier info) in the SciFinder record. This analysis excludes MEDLINE records, again without a warning message. If you are searching a set that has only CA references, I am quite certain that this analysis is identical to the Index Term analysis at least all the counts I reviewed are identical.

4) Supplementary Terms - Originally Supplementary Terms contained "single terms from the CA keyword phrases", which are [or were] indexing terms used to prepare quick indexes to each issue of printed CA. Keywords reflect the content of the title and the abstract, using vocabulary found in the original document. For CAPLUS, it analyzes the ST field. For MEDLINE, it analyzes the title field. Hence, MEDLINE records are not excluded from this analysis.

Use in Searching

1) CA Section Title - assuming I did not care about MEDLINE records in my set, this is handy when one needs a very broad category like Enzymes or Biochemical Genetics or Mammalian Hormones. This is useful when one wants categories too broad to be defined by keywords or to eliminate noise from a disparate category. It can also be useful for large sets where an index term analysis is overwhelming. Be sure to move far enough down in the Analyze listing to pick up older variant section titles for the same section. References often are assigned to more than one section, still care is needed since it is unreasonable to expect every single reference on enzymes regardless of the context to be in the Enzyme section. When I select CA Section Titles, I make the assumption I am selecting references where the major emphasis of the article (not unlike the asterisked headings in Medline) is related to the section.

2) Index Term - I use this most frequently and has the advantage of keeping MEDLINE records. In bibliographic instruction, I always say that we all get to a point in the search for where we have put in all the concepts we can think of and all the limits that we feel are safe, but still have too many references to comfortably browse. The Analysis by index terms is the perfect solution showing us what is in the set when we don't know exactly what we want. It generates ideas as to what facets of the set we want to look at.

3) CA Concept Headings - I have little use for this option since it basically performs an index term analysis. The only use I can think of is where I have a set of both CAPLUS and MEDLINE records and want to simultaneously eliminate the MEDLINE records while looking at the CA indexing.

4) Supplementary Terms - my bias is to strongly prefer index terms. Still, especially if going after a very new, specific, or unusual topic, it would be an extra precaution to check Supplementary Terms to make sure that an index term analysis had not missed some important records. Although I just found this out by researching this question, it does occur to me that this is a way to do a title term search in MEDLINE something that otherwise could only be done with the Explore References: Journal search screen. I think I would always first do an Analyze by Index Term first with ST as a double check, but seldom do an ST analysis alone.

Registry Number analyze is performed on additional indexing content provided by CAS and provides insights on the actual substances found in the literature being analyzed.

18b. Categorize is a powerful tool that tries to combine the best parts of index term/concept heading searching. Categorize as Analysis on steroids. In my opinion, it's best used on a somewhat smaller set, one that is being carefully defined. That is not to say that one cannot run Categorize on fairly large diffuse sets and make good use of it.

Still Categorize with its three-level analysis can be a little overwhelming. Though it is easy enough to use, I still think it is best utilized once users are willing to take a fair amount of time to really explore the rich depth inherent in this feature. In other words it's best when the patron really wants to take time looking at a set.

19. KMP (Keep Me Posted) <http://www.cas.org/support/scifi/etrain/kmp.html>

KMPs are intended to retrieve records when they are newly added to the database or modified (i.e. when the abstract or indexing is added). You can also set up the KMP profile so that duplicate answers are not retrieved by doing the following: Click on the 'Keep Me Posted Results' link, then click the Edit link beside the profile name. In the pop up window, you can check the box to 'exclude previously retrieved results'.

20. SciFindertalk

There is a listserv, SciFindertalk, which you could join, for current questions/discussion as well as to search the archives for discussion on any issues you run into.

<http://listserv.cas.org/SCRIPTS/WA-CAS.EXE?A0=SCIFINDERTALK>

21. VPN

There is a possible problem with the use Internet Explorer 6 and 7, with the IE security setting "Userdata persistence" disabled. The workaround is to enable "Userdata persistence" within IE security settings. Open Internet Explorer - select Tools > Internet Options - go to Security tab - select "Internet" zone - select "Custom level" - under Miscellaneous, find the "Userdata persistence" option and ensure that it is enabled.

22. Topically oriented learning scenarios

SciFinder - Part of the Chemical Synthesis Process

<http://www.cas.org/ASSETS/DB0D3AD423644A22874D00CCDBBF3D83/sfsynth.pdf>

SciFinder Tutorials – Web Version

<http://www.cas.org/support/scifi/tutorials.html#elearn> -- "e-Learning modules" under Substances and Reactions

SciFinder Scholar – Materials for Instructors

<http://www.cas.org/support/academic/sf/instructor.html> -- homework question examples

STNews – Database News

<http://www.cas.org/support/stngen/stnews/dbnews/index.html>

SciFinder Solutions - Techniques and Tips

<http://www.cas.org/support/scifi/sfsolutions/index.html>

23. Description of CA print (with a comparison with SciFinder)

<http://www.lib.utexas.edu/chem/info/ca.html#discardingca>

24. Markush searching ...

The Markush structure search feature **only** searches the MARPAT database, **not** the RegFile.

There are more than 800,000 searchable Markush structures from patents since 1961.

Alloys, metal oxides, inorganic salts, intermetallics, and polymers are not included.
Updated daily with approximately 60-75 patent citations and 150-200 Markush structures

Markush results are CAPlus patent literature records, **not** a set of structure records. The CAPlus patent literature records are equivalent to the corresponding Marpat patent literature record. Thus, neither the MARPAT records nor the Markush structures are displayed.

The CAPlus patent records may contain an embedded structure drawing in the abstract. This embedded graphic (which appeared in the print CA as well) often is expressed in Markush conventions. Alternatively, perusal of either the title, abstract, and keyword indexing or the full-text of the patent is required, since **no** CAS Registry Numbers are displayed.

Ring atoms and ring groups (Hy (any heterocycle), Cy (any cycle), Cb (any carbocycle) in query are matched only to real atoms in the Markush structures. For example, a query pyridine may match only a result pyridine, not a generic Hy.

Markush structures are a simple concept for chemists, but the conventions for the actual computer algorithms that map between a query structure and stored target Markush connection tables with a mix of generic and real atoms in both the query and the target structures is extremely complex.

25. Show precision analysis (should be the default setting).

'Show precision analysis' provides a breakdown of substructure search results (Conventional Substructure, Closely Associated Tautomers and Zwitterions, Loosely Associated Tautomers and Zwitterions, and Other). This is important for organometallics since metal atoms **cannot** be locked out to further substitution.

For example: Searching for 38833-34-0 as a substructure, with the Ru atoms either locked out of further substitution or not locked out of further substitution, and checking off Single component ... retrieves a mishmash of 915 structures.

Checking off 'Show precision analysis', however, with a locked structure search gives:

Conventional Substructure 90 ... apparently all the 'correct' structures
Closely Associated Tautomers and Zwitterions 819

Checking off 'Show precision analysis' for an unlocked structure search gives:

Conventional Substructure 135 (the additional substances have substitution on both Ru atoms)
Closely Associated Tautomers and Zwitterions 775

"The SciFinder structure search algorithm is designed to favor recall over precision, so there are rules in place that in some cases can cause manipulation of the user query, especially when the query contains metals:

Structures that contain metals are automatically manipulated to allow for greater variations. Disconnected and ring forms are retrieved regardless of how their structure is represented. All hydrates are automatically retrieved for metal-containing substances. Smart Search: <https://www.cas.org/help/scifinder/smrtsrch.htm>

By disregarding the lock, SciFinder will improve recall when searching for information regarding chelate bonds to metal centers, since, in solutions, metal bonding patterns may change, for instance, with some bonded atoms being replaced by coordinating solvents, and when authors are simply guessing at the bonding patterns for studied compounds

A by-product of this expanded recall can be, of course, the retrieval of answers that are less desirable. This is what led to the development of the precision analyze feature. Users that are just looking for a few good answers may use this feature to reduce the number of hits to study, but at the risk of missing some important literature depending on their needs.”

Note that if, a substance contains any stereochemical bonds, there is not an option of analyzing by precision using the Web version of SciFinder (although one can do a precision analysis at any point using the client).

26. [Help/Contents/Explore Tasks/ Draw Structures/Vertical Tool Palette/ Repeating Group Tool]

Draw a structure containing the group to be repeated.

Click [Repeating Group tool].

The cursor changes, and text boxes for: From/To are displayed.

Click a node, or nodes (to enclose the portion of the structure that you want as a repeating group).

To select multiple nodes, use [ctrl]

A rectangle encloses the area, which is then highlighted.

Exact and similarity searches: A single value (e.g., 5-5)

Substructure and reaction searches: Either a single value or range of values (0-20 is maximum)

27. [Help/Contents/Explore Tasks/Draw Structures/Vertical Tool Palette/ Variable Attachment Position Tool]

The Variable Attachment Position tool allows you to: Specify a node or fragment, and attach it to any of several positions on a ring system for searching. One or more substituents can be variably attached to the same nodes in the ring system.

Restrictions: No more than 20 substituents can be variably attached to ring systems. Variable attachment positions can be used only in substructure or reaction searches or when refining an answer set.

Draw the Ring System and the node(s) or fragment(s) to be attached.

Nodes/Fragments can be: atoms, shortcuts, variables, R-groups, substances, etc.

On the Vertical Tool palette, click [Variable Attachment Position tool]. The cursor changes.

Click the node or fragment and drag the cursor to the desired attachment point(s). A connecting line(s) displays.

28. CAS Roles, Predicted Properties and Experimental Properties

Substance Roles

<https://www.cas.org/help/scifinder/index.htm> - Index/Roles/substance

Experimental & Calculated Property Groups (Detail View)

<https://www.cas.org/help/scifinder/index.htm> -- Index/Properties/groups

=====