Searching Chemical Literature Using STN Easy

CHEM 355


Chemical Abstracts website: http://www.cas.org/

One does not have to study much chemistry before one requires information that is not contained in textbooks. Initially, the information needed is physical properties of compounds—melting points, boiling points, density, $pK_a$, etc. and can be found in chemical handbooks and catalogs. These reference materials usually only contain data on fairly simple compounds, those that are frequently used as student unknowns and/or starting materials in chemical synthesis. When one needs information that is beyond the scope of these sources, one must delve into the primary (journal articles, patents and dissertations) and secondary (review articles and books) literature. This can be an intimidating task, since there are literally thousands of journals that publish research of a chemical nature. Fortunately, most significant organic chemistry is confined to approximately one hundred of these!

Clearly, wandering the library stacks and browsing through journals is not the way to proceed. The researcher generally consults one of the major “guides” to the literature first. The most important of these are Chemical Abstracts, Beilstein, and Science Citation Index. Beilstein is most useful for locating information on a specific chemical compound. Science Citation Index is useful if one already knows the names of specific researchers in the field, or has a lead paper from which to build a search.

Chemical Abstracts Service (CAS) is a division of the American Chemical Society that produces the world’s largest and most comprehensive chemical information databases. CAS currently abstracts information from around 9000 journals and other document types and adds over 500,000 entries to their databases each year. In addition to producing hard-copy indexes, the CAS databases are accessible for searching via the Scientific and Technical Information Network (STN) and the WWW—for a fee (Beilstein is also searchable on-line). The two main databases are the Chemical Abstracts (CA) file and the Registry (REG) file.

Despite the fact that these electronic databases are only available for information appearing since 1967, the CA and REG files include over 14 million document abstracts and
over 20 million chemical substances. Over 4,500 new chemical substances are added to the REG file every day!

STN Easy is probably the simplest way to access the electronic CAS databases. The trade-off for this simplicity is some limitations on the types of searches one can do. STN Easy cannot search structures, for example. Furthermore, the use of Boolean operators has been partially supplanted by conducting searches with various menu options, although the searcher still has several useful options in the “Advanced Search” mode.

Remember that using STN to search CAS costs money! There are nominal fees for login and each search performed. **Most of the cost is incurred for the information you actually VIEW, however.** In the author’s opinion this is the most equitable way to charge for the service. In the past, the amount of time one spent logged-in was a major cost factor, which made efficiency of searching a top priority. The current system is friendlier to those who are not experts at the search process. The educational subscription to STN Easy restricts searching to the period from 5 PM to 4 AM (?) on Sunday-Friday and until 3 PM on Saturday. Searching during these “off-peak” hours saves the educational user 90% of the regular cost!

Things to remember regarding cost:

1. You can view “cost since login” at any time.
2. “Refining” a search is FREE. Thus one should make the original inquiry fairly broad, and narrow the hits down through the no-cost refining process to pinpoint the information one needs.
3. Since most of the cost is for the hits that are displayed, one should only display a hit when fairly certain it will contain useful information. Displaying every hit from a search (if it is more than a dozen or so) can be wasteful.
4. Unless you (or a colleague) can read languages other than English, you may want to restrict the hits you view to those where the original article is in English. This is one of the search and refining options.

Before sitting down to the terminal to login to the database (a login ID and password are required), one needs to go through a few preliminary steps. First, one should define the goals of the search: Are you trying to determine if a substance has been reported in the literature? Are you looking for papers written by a particular person? Maybe you need a method to prepare a particular functional group or compound. Next, determine what databases and/or files are likely to contain the information you need. For example the REG file (“chemical substances” in STN Easy) will be able to tell you if a particular compound has been reported in the literature, its CAS name (sometimes different than IUPAC!), and its CAS registry number. **The registry number is essential for finding out more about the compound.** The CA file (“chemistry references” in STN Easy) contains the abstracts for the original literature. This is the file to search for authors, keywords, and synthetic preparations.
The following is a brief, general set of instructions for using STN Easy. Even though some general protocols are always followed, each search is different. Blindly following a set of instructions without analyzing the “hits” from each inquiry will lead to inefficiency and missed information. The instructions start from the point just after log-in (which must be done by a faculty member). It will probably be necessary to get assistance with your first search attempt to reduce questions, frustrations, and cost.

**Searching for a CAS registry number**

1. On the Basic Search page choose **Chemical Substances** as the **Category**, then go to **Advanced Search**.
2. Under the Advanced Search option, choose **Molecular Formula** as the search field and enter the molecular formula. The formula must be in the format C\textsubscript{x}H\textsubscript{y}A\textsubscript{z}BCD… where \(x\) is the number of carbons, \(y\) is the number of hydrogens, and ABCD… are the remaining elements in alphabetical order. For example, 3-chloro propanol [ClCH\textsubscript{2}CH\textsubscript{2}CH\textsubscript{2}OH] would be entered as C\textsubscript{3}H\textsubscript{7}ClO.
3. When the formula has been correctly entered, click on **Search**.
4. This search will likely produce a hit list from one to thousands of compounds. Time to refine the search!
5. Click on **Refine**. It is usually best to refine the list using the complete or partial name of the compound you are looking for. It is very time consuming to generate the complete (and correct!) chemical name, however. The author finds it more convenient to refine the search looking for **Word(s)** that are a portion of the chemical name. The **Refine** screen has an operator field and a search field.
6. When you have refined in the above manner until one to few dozen hits are left, the hits will be displayed with partial chemical names. Scan down the list until you see a compound that may be the one you want. Click on the hit to display it. If it is in fact the molecule you were seeking (check it against the displayed structure) copy the registry number from the entry or print that screen. If the structure displayed is NOT the one you were seeking, you can still learn from that entry. Look at how the compound is named. Did you try to refine with an incorrect word or name fragment? If so, go back and try the refinement again. If you still think your name fragment is OK, continue scanning down the list of hits looking for the desired name.

**Searching for references using the CAS registry number**

Frequently when you display the page containing the structure and registry number for the compound of interest, a link above the structure says “Information related to CAS RN XXX-YY-Z is available in other categories” of the CAS files. A window at the bottom of the page shows the possible restrictions for further searching on the registry number. There is no
guarantee that references to the compound are present in all of the categories. Here you can restrict searching of the Chemistry References using any or all of these categories.

Alternatively, you can follow the basic set of instructions below.

1. Select a New Search and choose Chemistry References as the category.
2. Enter the CAS registry number in the search field and click on search. (Remember you can refine the hit list for free later.)
3. Refine the hit list as your need dictates. Possible restrictions for refinement are document type, author, year of publication, etc. This is the time to refine for a synthetic preparation or application. The most likely sources to contain information about how to make a compound and its physical properties would be of this type.
4. Display and/or print only those hits you think are likely to contain the information you need. Keep in mind that the journal article you find using STN may not have the information, but may list one or more footnotes that leads you to an earlier paper that does have the information.

The Chemistry References database is also the place to search keywords, author names, etc. Once you have your lead references from the STN Easy search, it is time to go to the library stacks and sleuth out the information you need.
Instructions for “The Aldol Condensation Reaction: Preparation of Benzalacetones and Benzalacetophenones”

Literature Search Using STN Easy

**Variation**: Chalcones from various substituted benzaldehydes and acetophenones:

1. Login to STN Easy from the CAS web page: URL http://stneasy.cas.org/ (this will require the assistance of a faculty member).
2. Under **Select Category** choose **Chemical Substances** and **Go to Advanced Search**.
3. Check the **cost since login** and jot the dollar amount down. Since some of your classmates may have been using STN Easy in the same session without logging out of STN, the amount may be substantial—do not panic.
4. In the **Advanced Search** screen, choose molecular formula from the search field choices and enter the molecular formula of your assigned chalcone. Remember that the formula must be in the format: carbon, hydrogen, other elements in alphabetical order. Click on **Search**.
5. How many answers to that search question were there? **Refine** the search question using **AND word(s)** “2-propen-1-one” All chalcones have the same parent name: (2-propen-1-one).
6. Now how many answers are in the hit list? If there are still too many to easily scan through manually, **Refine** the search further with additional modifiers appropriate for your compound (for example the **word(s)** 4-nitrophenyl). Remember refining a search does not cost anything!
7. Scan the refined answer list looking for a name that fits your assigned compound. When you find your compound, print the screen containing the structure and CAS registry number. Remember to double-check the structure—are the substituents in the correct place? is the double bond the correct configuration (E or Z)?
8. Is information related to your compound available? Start by looking at the complete list of references. How many are there? **Refine** the list to just references since 1980—how many does that include? Record these answers for your lab report.
9. Now search the various categories one at a time for specific references to your chalcone. Try **no restrictions** first. If there are too many references, then **refine** the list, but be sure to check at least **synthetic preparation** and **reactant** and **therapeutic use**. Refine the lists further where necessary—to include only journals articles, for example. Many of the “good” chalcone papers are in languages other than English, so don’t refine the language. When you have refined the lists of references to a handful of answers, begin to display the entries. Pay particular attention to the source—the journal in which the original article appeared. If it is a
journal contained in the WWU library, or one that is commonly available by interlibrary loan (ILL) print that reference for future use.

10. When you have finished gathering references about your assigned chalcone, check the cost since login, and determine how much the information cost at regular prices. Record the amount for the lab report form.

11. Look up the references you have found in the library (or by ILL) and find the physical properties, methods of preparation, and spectral data for your assigned compound. Remember that since the CAS online starts in 1967 older references will not be found directly. You may have to use references in the papers you found to access older literature for the information you need.

12. A search of Beilstein should also be conducted on your chalcone.