

Making use of Substance Records in Reaxys

Content found in Substance Records and how to find it

Learning Objectives

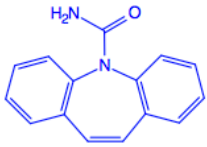
- Understand how information in Substance Records is structured and organized
- Differentiate between property fields, subfields, and data points
- See where Substance Records contain text terms and/or numeric terms
- Master different search strategies used to find different types of information in Substance Records
- Learn how to use the Substance Basic Index.

Content and Search Opportunities in Substance Records

Indexed substances are catalogued in over 28 million Substance Records, after carefully selection from approximately 450 major chemistry and chemistry-related journal titles and key patent sources. The selection focuses on the most important substances in the original documents, and these are excerpted along with factual information about them.

A Substance Record contains two basic types of information:

- A. **Substance identifying information**, such as structure, name, and formula.



NC(=O)c1ccc2c(c1)ccc3ccccc23

carbamazepin

Identification

Physical Data - 308

Spectra - 123

Bioactivity - 1286

Other Data - 1460

Preparations - 54 >

Reactions - 162 >

Documents - 2478 >

carbamazepin ×

| | | | |
|-------------------------|---|---------------------------|--|
| Reaxys ID: | 1246090 | | |
| Chemical Names: | carbamazepin, 5H-dibenz(b,f)azepine-5-carboxamide | | |
| CAS Registry Number(s): | 298-46-4 | Substance type: | |
| Molecular Formula: | C ₁₅ H ₁₂ N ₂ O | Linear Structure Formula: | C ₁₅ H ₁₂ N ₂ O |
| Molecular Weight: | 236.273 | No of references: | 2478 |
| InChIKey: | FFGPTBGBLSHEPO-UHFFFAOYSA-N | | |

Fig 1. Substance identifying information for carbamazepine

B. **Substance property information**, which is organized primarily into property fields and subfields. Text or numeric information in a given subfield is called a data point.

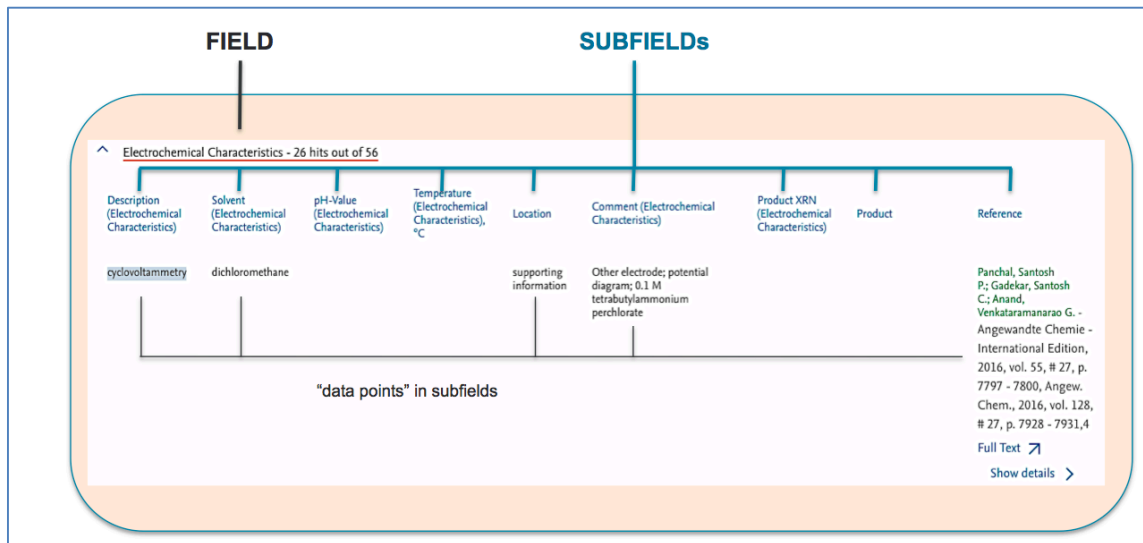


Fig 2. Example of fields, subfields and data points

Property information of a substance can be displayed by:

- finding the substance of interest and
- then clicking the various links that lead to property information (e.g., **Physical Data** in Fig. 1)

Property information can also be searched:

- some fields/subfields via **Quick search**
- most fields/subfields via **Query builder**

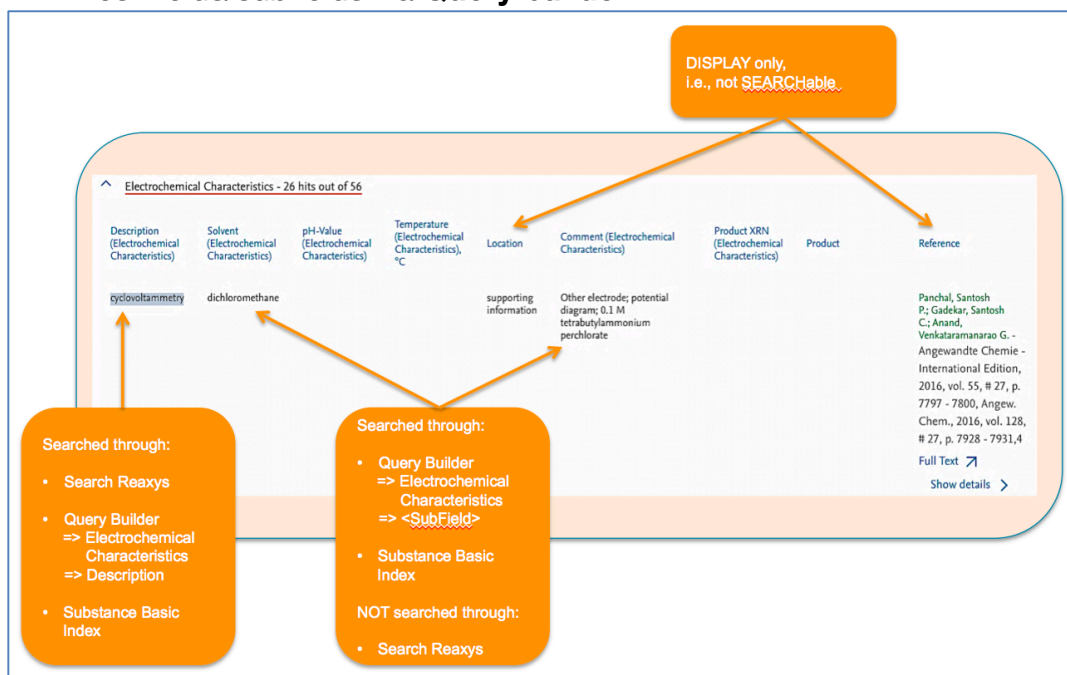


Fig 3. A summary of property search opportunities

The Substance Basic Index: a special search field

As we become familiar with the property fields and subfields in Reaxys, we can easily find information via **Quick search** by entering and searching terms found in the Description subfield of a property.

But what if we don't know the relevant term or in which subfield it would appear? How do we then proceed? We shall learn how to solve this in the examples below, but first it helps to know that there is a special search field, the **Substance Basic Index Field**, which searches text terms in all fields of Substance Records. Some important points to note:

- only single terms (not phrases) are searched via Substance Basic Index.
To search *lipase inhibitors*, we would need to connect the terms with operators (e.g., PROXIMITY or NEAR) to retrieve only Substance Records where the terms are close to each other
if we only enter *lipase inhibitors*, Reaxys automatically connects the two terms with the operator AND. This means that *lipase* could appear in one subfield and *inhibitors* in another, so answers in the hit set would not be very precise.
- we can use truncation (*) to search for all terms with the same word stem.
- Data points in some subfields are not searchable, e.g., the **Location** and **Reference** subfields in Fig 3.

Some examples to try

By working through the following examples, we can learn how different searches 'work'. Make note of the results in each example, and consider which workflows may be used to solve some of your own search questions.

Substance Identification

1. Search the following terms using the **Search Reaxys** entry box of **Quick search**. Comment on whether results are the substance in Fig.1 alone, with other substances, or not at all.
 - a. carbamazepin
 - b. tegretol
 - c. 298-46-4
 - d. C₁₅H₁₂N₂O
 - e. 236.273
 - f. FFGPTBGBLSHEPO-UHFFFAOYSA-N

Main learning point: There are several ways to find individual substance records through **Quick search** and Reaxys often shows only some of the names it recognizes for a substance (Reaxys has an independent substance name dictionary). Additionally, when Reaxys recognizes the name of a substance that has a structure connection table, it retrieves hits with the same connection table (i.e., stereoisomers, isotopic substances).

This example is one of the rare occasions where there are multiple registrations in Reaxys for the same substance (Reaxys is progressively attending to these rare occurrences).

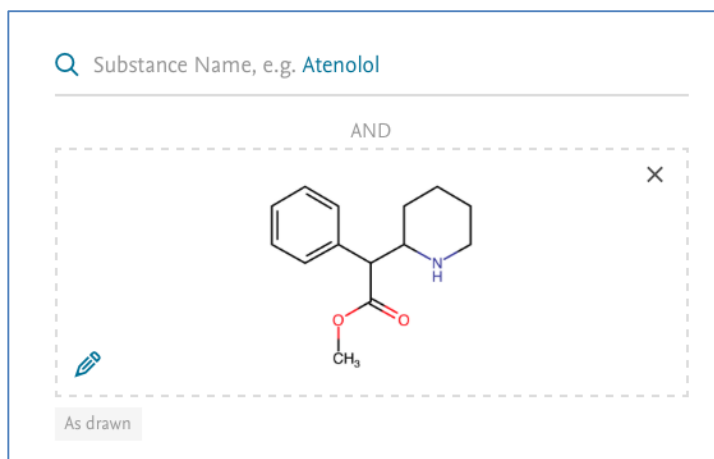
2. Perform a text search for the substance Cs_2SO_4 in **Quick search** and comment on the substances you retrieve.

Main learning point: Often, inorganic substances can be searched by typing the formula into **Quick search** and Reaxys retrieves different crystalline forms.

3. Find *fluticasone propionate* via a text search in **Quick search**. Compare the Substance Records listed as numbers 1,2, and 4, checking for differences in entries like the structure diagrams or the InChIKey(s).

Main learning point: Reaxys enters stereoisomers exactly as they are represented in the original source documents. Nevertheless, a number of normalizations are done upon entering data into Reaxys, such as structure abbreviations (OEt to OCH_2CH_3), donor bonds (nitro groups are represented in standard structure conventions with the donor bond changed to a double bond) and so forth.

4. Comment on the different answer sets obtained when you perform:
 - a. a text search for *methylphenidate* in **Quick search** versus
 - b. a structure search for the same (below). Generate the query in the structure search using the function **Create structure template from name** (don't change any defaults):



Main learning point: Most users are interested mainly in exact searches for substances, but there are occasions where salts and mixtures in particular are also required. This exercise shows how the different answer sets are easily obtained.

Substance Basic Index (SBI)

Add the **Substance Basic Index** to the workspace of Query builder by dragging it from the menu on the right. You will find it listed under **Basic Indexes**. Make sure to specify **Substances** when you click **Search**.

In each of the following examples, examine the **Hit Data** (there may be several of them for a substance) and determine why the answer(s) was retrieved.

5. Enter *carbamazepine* into the **Substance Basic Index** and search.
 - a. How many Substance Records did you retrieve?
 - b. Comment on the default ranking (No of References).
 - c. Is the Substance Record for *carbamazepine* in the answer set?

[Hint: **Filter** the hit set by **Structure** and make sure to click **Limit to** after transferring the structure]

Main learning point: The SBI searches single words in all property fields and subfields, including the **Chemical Name** field (but not in the display only subfields shown in Fig 3).

6. Enter *lipase inhibitor** into the **Substance Basic Index** and search.
 - a. How many Substance Records did you retrieve?
 - b. Browse through some of the records to see which fields/subfields contain the Hit Terms.

[Hint: the **Hit Data** in the Substance Record for nitrogen may illustrate the issue most quickly.]

Main learning point: If querying more than one term, we always need to consider how to connect these. Reaxys defaults to AND, but at times this may not be precise enough.

7. Perform the following searches in the **Substance Basic Index** and compare results:
 - a. lipase proximity inhibitor*
 - b. lipase near inhibitor*
 - c. lipase near inhibitor

Main learning point: In the SBI it is best to connect terms with either the NEAR or PROXIMITY operators; it often helps to try both. Also, always consider the impact of truncation. In this case, terms included in the search through *inhibitor** are *inhibitor*, *inhibitors*, *inhibitory*. We should use truncation carefully to increase comprehension of answers but still maintain reasonable precision.

8. In some of the hits for example 7, we find *pancreatic lipase of porcine* as **Test System**. What search in the **Substance Basic Index** would we do to retrieve with reasonable precision substances tested on porcine pancreatic lipase? Try a few options before deciding on a final answer.

9. There are a few ways to find which fields and subfields in Reaxys may contain information of interest. Try these searches:
 - a. Query *cyclovoltammetry* in **Quick search**. How many answers do we get, and how were they obtained? Also, look for systematic terms for this technique in some of the **Hit Data**. We can use these terms subsequently in **Quick search** to obtain information in the subfield.
 - b. Enter *cyclovoltammetry* into the **Substance Basic Index**. Look for systematic terms for this technique in some of the **Hit Data**. We can use these terms subsequently in **Quick search** to obtain information in the subfield.

Main learning point: We can find systematic terms for properties in either way described above (the next example, shows a third way).

Text searches for properties in Query builder

10. Another way to find systematic terms for properties in Substance Records is to use the Lookup function in **Query builder**. Go to **Query builder** and find the field **Electrochemical Characteristics** in the menu on the right (use the **Search Properties** box!). Drag the field into the workspace and display its subfields. Then:
 - a. Click the **Lookup** icon of the subfield **Description (Electrochemical Characteristics)** and note entries on each of the 2 pages that appear in the dialog box.
 - b. Check the boxes: *cyclic voltammetry* and *cyclovoltammetry*, and then click **Transfer**.

Note: Index terms may change with time, but Reaxys automatically searches older/newer index terms in property fields.
 - c. Click **Search (Substances)**.

Main learning point: Knowing systematic terms always helps improve search strategies. The Lookup function in a relevant **Description** subfield is an easy way to find systematic terms.

11. Find the list of solvents in which various electrochemical studies recorded in Reaxys have been measured.

[Hint: Look up terms in the **Solvent (Electrochemical Characteristics)** subfield.]

Main learning point: **Lookup** in most subfields can help find terms to use in our search strategy. When the resulting list of terms is too long to scroll through, search potential terms of interest using the **Search** tool at the top of the dialog box.

Numeric searches for properties in Query builder

Numeric searches usually need to be done in a numeric subfield of a relevant field, which can be found through **Query builder**.

12. Talking to a colleague in the School of Physics, she mentions to us that she is researching the Curie temperature of materials. She adds there are a few values mentioned in Wikipedia and some tables of Curie temperatures on the Internet. However, she wants to find substances with recorded Curie temperatures at room temperature. We suggest: “Have you tried Reaxys?” She replies: “Are you crazy?” Demonstrate that we are not crazy. Find Substance Records in Reaxys with recorded Curie temperatures at room temperature (20–35°C).

[Hint: Use any of the methods above to find the field that systematically lists the property Curie temperature. Then, enter *Curie temperature* in the **Description** subfield and 20–35 in the **Temperature** subfield (temperature is listed in °C).

Main learning point: Many numeric fields in Reaxys have thousands of numeric values, which are important in a range of disciplines, like Physical Chemistry, Physics, and Materials and Engineering Sciences.

13. In our laboratory, we have isolated a natural product that has:
- A melting point of 200–202°C
 - a molecular ion in the high-resolution mass spectrum at 645.737 (for those not familiar with this technique, then the molecular ion is the precise Molecular Weight)
 - An optical rotatory power of +18° at 589 nm.

We also know that the substance is biologically active and it has recently been reported to have a neuroregulatoric effect.

What is the substance, from what plant family is it commonly isolated, and in what year was the reported neuroregulatoric effect published?

[Hint: When searching numeric values, consider searching within a range of values because different source documents will likely report slightly different values. So, in this case perhaps search a melting point range of 198–204, a molecular weight range of 645–646, and an optical rotation range of 17–19 (at 589 nm). This way, we should be able to find the substance, but it has quite a lot of pharmacological data. How might you search for the effect using the SBI?]