

## 4.1: PubChem Web Interfaces for Text

### PubChem Homepage

The PubChem homepage (<https://pubchem.ncbi.nlm.nih.gov>) provides a search interface that allow users to perform any term/keyword/identifier search against all three major databases of PubChem<sup>1,2,3</sup>: Compound, Substance, BioAssay. If a search returns multiple hits, they are presented on an Entrez DocSum page and will be explained in more detail later in this chapter. If the search returns a single record, the user will be directed to the web page that presents information on that record. This page is called the [Compound Summary](#), [Substance Record](#), or [BioAssay Record](#) page, depending on the record type (i.e., compound, substance, or assay). In addition, the PubChem homepage provides launch points to various PubChem services, tools, help documents, and more. In general, the PubChem homepage is a central location for all PubChem services.

### Entrez Search and Retrieval System

NCBI's Entrez<sup>4,5,6,7</sup> is a database retrieval system that integrates PubChem's three major databases as well as other NCBI's major databases, including [PubMed](#), [Nucleotide](#) and [Protein Sequences](#), [Protein Structures](#), [Genome](#), [Taxonomy](#), [BioSystems](#), [Gene Expression Omnibus](#) (GEO) and many others. Entrez provides users with an integrated view of biomedical data and their relationships. This section focuses on search and retrieval of PubChem data using the Entrez system. A more detailed description on the Entrez system is given in the following documents:

- The Entrez Search and Retrieval System  
(<http://www.ncbi.nlm.nih.gov/books/NBK184582/>)
- Entrez Help  
(<https://www.ncbi.nlm.nih.gov/books/NBK3836/>)

### Entry points to Entrez

One can search the PubChem databases through Entrez, by initiating a search from the NCBI home page (<http://www.ncbi.nlm.nih.gov>). By default, if a specific database is not selected in the search menu, Entrez searches all Entrez databases available, and lists the number of records in each database that are returned for this “global query”. The following link directs you to the global query result page for the term “AIDS” against all databases integrated in the Entrez system.

<https://www.ncbi.nlm.nih.gov/gquery/?term=AIDS>

Simply by selecting one of the three PubChem databases from the global query results page (under the Chemical section), one can see the query results specific to that database.

Alternatively, one can start from the PubChem home page (<http://pubchem.ncbi.nlm.nih.gov>), where a search of one of the three PubChem databases may be initiated through the search box at the top. It is also possible to initiate an Entrez search against a PubChem database from the following pages:

- <https://www.ncbi.nlm.nih.gov/pccompound/> (to search the Compound database)
- <https://www.ncbi.nlm.nih.gov/pcsubstance/> (to search the Substance database)
- <https://www.ncbi.nlm.nih.gov/pcassay/> (to search the BioAssay database)

### Entrez DocSums

If an Entrez search for a query against any of the three PubChem databases returns a single record, the user will be directed to the Compound Summary, Substance Record, or BioAssay Record page for that record (depending on whether the record is a compound, substance, or assay). If it returns multiple records, Entrez will display a document summary report (also called “DocSum” page). The following link directs you to the DocSum page for a search for the term “lipitor” against the PubChem Compound database:

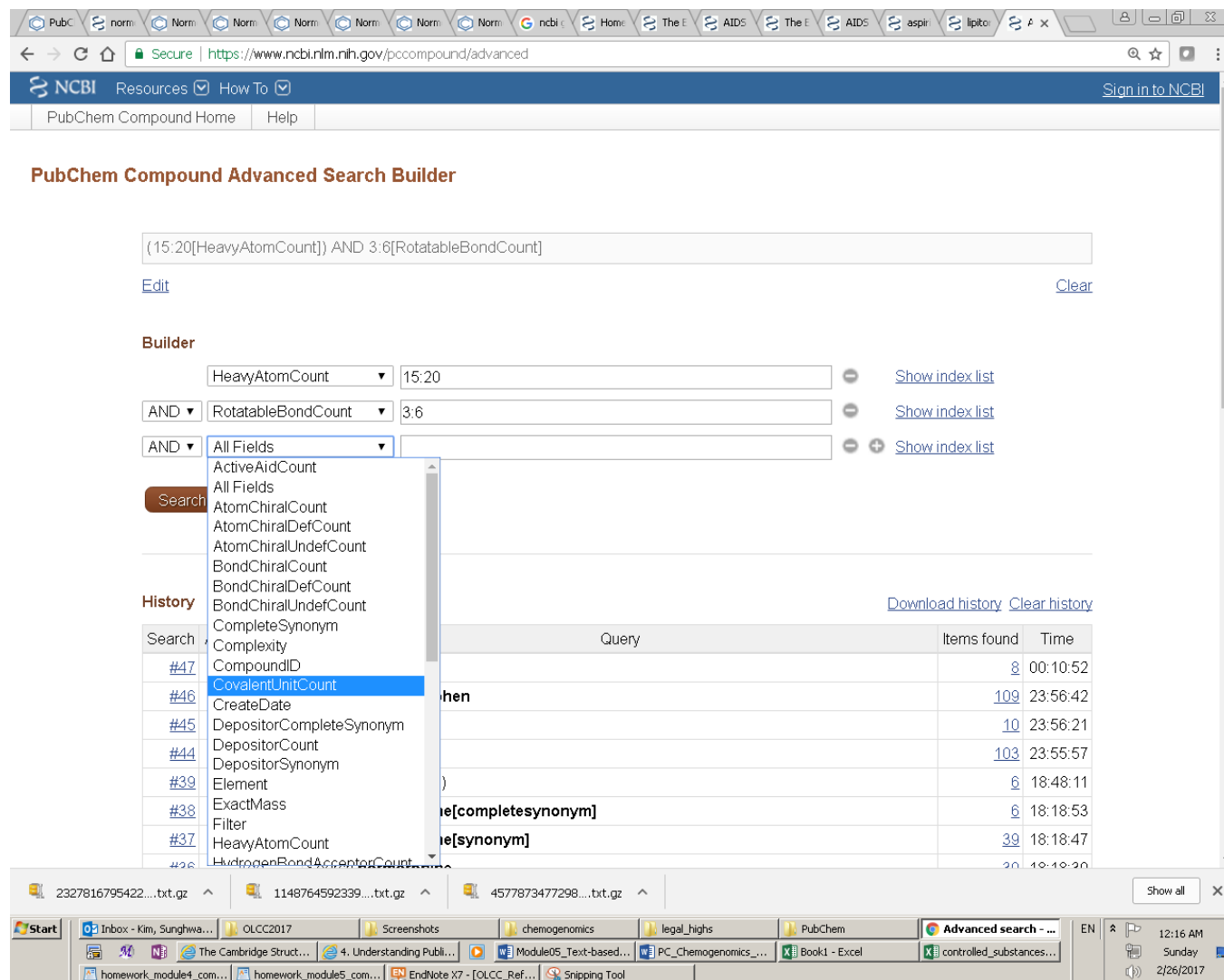
<https://www.ncbi.nlm.nih.gov/pccompound?term=lipitor>

In this example, the DocSum page displays a list of the compound records returned from the search. For each record, some data-specific information is provided with a link to the summary page for that record. The DocSum page contains controls to change the display type, to sort the results by various means, or to export the page to a file or printer. Additional controls that operate on a

query result list are available on the right column of the DocSum page. The DocSum page for the other two PubChem databases look similar to this example for the Compound database.

## Entrez Indices

Entrez indices, tied to individual records in an Entrez database, include information on particular aspects (often referred to as fields) of the records. These indices may have text, numeric or date values, and some indices may have multiple values for each record. The available fields and their indexed terms in any Entrez database can be found from the drop-down menus on the Advanced Search Builder page (which can be accessed by clicking the “Advanced” link next to the “Go” button on the [PubChem Home page](#)).



PubChem Compound Advanced Search Builder

Search query: (15:20[HeavyAtomCount]) AND 3:6[RotatableBondCount]

Builder:

- HeavyAtomCount: 15:20 [Show index list](#)
- AND RotatableBondCount: 3:6 [Show index list](#)
- AND All Fields [Show index list](#)

History:

Search	Query	Items found	Time
#47	hen	8	00:10:52
#46		109	23:56:42
#45		10	23:56:21
#44		103	23:55:57
#39	)	6	18:48:11
#38	re[completesynonym]	6	18:18:53
#37	re[synonym]	39	18:18:47
#26		20	18:18:20

Download history [Clear history](#)

When the user enters a query in the Entrez search interface, the Entrez indices are matched directly to that query. By default, in an Entrez search with a simple query, all indexed fields are matched against the query, usually resulting in the largest number of returned records including many unwanted results. One can narrow the search to a particular indexed field, by adding the index name in brackets after the term itself (e.g., “**lipitor[synonym]**”). For numeric indices, a search for a range of values can be done by using minimum and maximum values separated by a colon and followed by the bracketed index name (e.g., “**100:105[MolecularWeight]**”). Multiple indices may be searched simultaneously using Entrez’s Boolean operators (e.g., “**AND**”, “**OR**” and “**NOT**”).

A complete list of the Entrez indices available for the three PubChem databases can be retrieved in the XML format, using the eInfo functionality in [E-Utilities](#) (which will be covered in Module 7):

- <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi?db=pccompound> (for Compound)
- <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi?db=pcsubstance> (for Substance)
- <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi?db=pcassay> (for BioAssay).

Additional information on the PubChem Entrez indices is available in the “Indices and Filters in Entrez” section of the help documentation:

[https://pubchem.ncbi.nlm.nih.gov/help.html#PubChem\\_Index](https://pubchem.ncbi.nlm.nih.gov/help.html#PubChem_Index)

## Entrez Links

Entrez links are cross links or associations between records in different Entrez databases, or within the same database. These links may be applied to an entire search result list (via the “find related data” section at the right column of a DocSum page) or to an individual record (via links at the bottom of each record presented on the DocSum page). The Entrez links provide a way to discover relevant information in other Entrez databases based on a user’s specific interests. Equivalently, one may think of this as a way to transform an identifier list from one database to another based on a particular criterion. Note that there are limits to how many records may be used as input in a link operation. To process a large amount of input records and/or to expect a large amount of output records associated with the input records, one should use the FLink tool (<https://www.ncbi.nlm.nih.gov/Structure/flink/flink.cgi>).

A complete list of the Entrez links available for the three PubChem databases can be retrieved in the XML format through these links

- <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi?db=pccompound> (for Compound)
- <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi?db=pcsubstance> (for Substance)
- <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi?db=pcassay> (for BioAssay).

## Entrez Filters

Entrez filters are essentially Boolean bits (true or false) for all records in a database that indicate whether or not a given record has a particular property. The Entrez filters may be used to subset other Entrez searches according to this property, by adding the filter to the query string.

Entrez filters are closely related to links in that the majority of Entrez filters in the PubChem databases are generated automatically based on whether PubChem records have Entrez links to a given database. However, some special filters, such as the “lipinski rule of 5” filter, or the “all” filter, are not link-based.

The Entrez filters available for each Entrez database may be found on the Advanced Search Builder page by selecting “Filter” from the “All Fields” dropdown and clicking “Show index list”.

## PubChem Compound Advanced Search Builder

[Edit](#)
[Clear](#)

**Builder**

Filter

all (94453557)  
has 3d conformer (83874522)  
has dailymed (5222)  
has mesh (119178)  
has no parent (646631)  
has parent (93806926)  
has patent (19483404)  
**has pharm (15417)**  
has src nih mlp (1012676)  
has src vendor (64531046)

Previous 200  
Next 200  
Refresh index

AND
All Fields

or [Add to history](#)

---

**History**
[Download history](#)
[Clear history](#)

Search	Add to builder	Query	Items found	Time
#47	<a href="#">Add</a>	Search <b>lipitor</b>	8	00:10:52

More detailed description of the Entrez filters available for the three PubChem databases are given in the “Indices and Filters in Entrez” section of the help documentation:

[https://pubchem.ncbi.nlm.nih.gov/help.html#PubChem\\_Index](https://pubchem.ncbi.nlm.nih.gov/help.html#PubChem_Index)


## Entrez History

Entrez has a history mechanism (Entrez history) that automatically keeps track of a user’s searches, temporarily caches them (for eight hours), and allows one to combine search result sets with Boolean logic (i.e., “AND”, “OR”, and “NOT”). The Entrez history allows one to limit a search to a subset of records returned from a previous search. Use of Entrez history can help users avoid sending and receiving (potentially) very large lists of identifiers. In addition, through the Entrez history, one can use the search results as an input to various PubChem tools for further manipulation and analysis.

## References

- (1) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L. Y.; He, J. E.; He, S. Q.; Shoemaker, B. A.; Wang, J. Y.; Yu, B.; Zhang, J.; Bryant, S. H. *Nucleic Acids Res.* **2016**, *44*, D1202.
- (2) Wang, Y.; Bryant, S. H.; Cheng, T.; Wang, J.; Gindulyte, A.; Shoemaker, B. A.; Thiessen, P. A.; He, S.; Zhang, J. *Nucleic Acids Res.* **2017**, *45*, D955.
- (3) Kim, S. *Expert Opinion on Drug Discovery* **2016**, *11*, 843.
- (4) Schuler, G. D.; Epstein, J. A.; Ohkawa, H.; Kans, J. A. *Methods Enzymol.* **1996**, *266*, 141.
- (5) McEntyre, J. *Trends in genetics : TIG* **1998**, *14*, 39.
- (6) The Entrez Search and Retrieval System (<https://www.ncbi.nlm.nih.gov/books/NBK184582/>) (Accessed on.
- (7) Entrez Help (<https://www.ncbi.nlm.nih.gov/books/NBK3836/>) (Accessed on.

4.1: PubChem Web Interfaces for Text is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.



4.1.4

<https://chem.libretexts.org/@go/page/170159>