

Module 5: How to search PubChem for chemical information (part 1)

Learning Objectives

- Explain what Entrez indices, filters, and links are.
- Explain what depositor-supplied and MeSH synonyms in PubChem are.
- Retrieve compounds that have a particular type of information (e.g., boiling point, melting point, and so on).
- Submit multiple text queries using the Identifier Exchange Service.
- Retrieve annotated information contributed by a given data source.
- Combine multiple queries using Entrez history.

1. PubChem web interfaces for text search

1.1. 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 PubChem¹⁻³: Compound, Substance, BioAssay (see Section 3 of [Module 4](#) for data organization in PubChem). If a search returns multiple hits, they are presented on an Entrez DocSum page, which was briefly mentioned in the [Module 4](#) Questions and will also be explained in more detail later in this Module. 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.

1.2. Entrez Search and Retrieval System

NCBI's Entrez⁴⁻⁷ 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/>)

1.2.1. 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)

1.2.2. 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.

1.2.3. 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

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

Edit Clear

Builder

HeavyAtomCount 15:20 Show index list

AND RotatableBondCount 3:6 Show index list

AND All Fields Show index list

Search

History

Search	Query	Items found	Time
#47	CompoundID	8	00:10:52
#46	CovalentUnitCount	109	23:56:42
#45	CreateDate	10	23:56:21
#44	DepositorCompleteSynonym	103	23:55:57
#39	DepositorSynonym	6	18:48:11
#38	Element	6	18:18:53
	ExactMass		
	Filter		

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

1.2.4. 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).

1.2.5. 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

"has pharm"[Filter] [Edit](#) [Clear](#)

Builder

Filter "has pharm"[Filter] [Hide index list](#)

- 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 [Show index list](#)

[Search](#) or [Add to history](#)

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

Search	Add to builder	Query	Items found	Time
#47	Add	Search lipitor	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

1.2.6. 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.

2. Text search in PubChem

2.1. Basics

Text search allows one to find chemical structures using one or more textual keywords, which may be chemical names (e.g., “aspirin”) or any word or phrase that describe molecules of interest (e.g., “cyclooxygenase inhibitors”). One can perform a text search from the [PubChem homepage](#), by providing a text query in the search box. If the query is **a phrase or a name with non-alphanumeric characters, double quotes should be used around the query**. Various indices can be individually searched by suffixing a text query with an appropriate index enclosed by square brackets (for example, the query “*N*-(4-hydroxyphenyl)acetamide”[iupacname]). Numeric range searches of appropriate index fields can be performed using a “:” delimiter (for example, the query 100.5:200[molecularweight] for a molecular weight range search between 100.5 and 200.0 g/mol). One can see what search indices are available in PubChem from the drop-down menu on the “[PubChem Compound Advanced Search Builder](#)”, which can be accessed by clicking the “advanced” link (next to the “Go” button) on the [PubChem homepage](#). Queries may be combined using the Boolean operators “AND”, “OR”, and “NOT”. These Boolean operators must be capitalized.

2.2. Depositor-supplied synonyms

Conceptually, data in a database are stored in the same way as we would record them in a table or excel spreadsheet. The rows in the table correspond to compounds, and the columns correspond to properties or descriptions for those compounds (e.g., melting and boiling points, chemical names, toxicity, bioactivity, target proteins, and so on). These columns are commonly called “data fields”. You may want to perform a search against all data fields or only a particular field. To search the (depositor-provided) chemical name field of the records in the PubChem Compound database, a chemical name query needs to be suffixed with either of the “[synonym]” or “[completesynonym]” index. The “[synonym]” index invokes search for molecules whose names contain the query chemical name as a part (that is, **partial matching**), and the “[completesynonym]” index invokes search for those whose names completely match the query (that is, **exact matching**). If no index is given after the query, PubChem will search all data fields. Compare the following searches for “aspirin” against the PubChem Compound database.

- **aspirin[completesynonym] (1 hit, as of Feb. 26, 2017)**
<https://www.ncbi.nlm.nih.gov/pccompound/?term=aspirin%5Bcompletesynonym%5D>
- **aspirin[synonym] (98 hits)**
<https://www.ncbi.nlm.nih.gov/pccompound/?term=aspirin%5Bsynonym%5D>
- **aspirin (103 hits)**
<https://www.ncbi.nlm.nih.gov/pccompound/?term=aspirin>

Note that the URLs for these searches contain the query strings (following the string “?term=”), and that the square brackets enclosing the Entrez indices “completesynonym” and “synonym” are replaced with the strings “%5B” and “%5D”. Because the first query resulted in only one hit, the user is directed to the Compound Summary page for the hit compound (CID 2244). On the other hand, because the other two queries result in multiple hits, the results are presented on the DocSum pages.

When either “[completesynonym]” or “[synonym]” is used, it is the “**depositor-provided synonyms**” fields of the compound records in PubChem that is searched for the query string. The depositor-provided synonyms field for a compound contains a filtered list of chemical names (synonyms) provided by individual data providers for the substances associated with that compound. These synonyms are presented in the “Depositor-provided synonyms” section on a Compound Summary page. To see the variety of synonyms for a compound, check the following link [to the Depositor-provided synonyms” section of the Compound Summary page for CID 2244 (aspirin)]:

<https://pubchem.ncbi.nlm.nih.gov/compound/2244#section=Depositor-Supplied-Synonyms>

For CID 2244, there are more than 700 depositor-supplied synonyms. These synonyms include not only those commonly used in chemistry class (e.g., common names, IUPAC names, CAS registry numbers) but also those used in many other places (e.g., database identifiers, chemical vendor catalogues, the name of products that contains the chemical, code numbers internally used in a company, and so on).

As mentioned above, the search for aspirin with the “[completesynonym]” index specified returns only one compound (CID 2244). It means that one of many names of this compound exactly matches the query string “aspirin”. On the other hand, the search for aspirin with the “[synonym]” index returns additional 97 compounds. It means that at least one of the names of each these compound partially match the query string (that is, the compound contains the string “aspirin” in one of its names). Interestingly, the results from the last two queries include acetaminophen (CID 1983), which is the active ingredient of Tylenol. Check the following link to the depositor-provided synonyms section of CID 1983 to see what synonyms of Tylenol contains the string “aspirin”:

<https://pubchem.ncbi.nlm.nih.gov/compound/1983#section=Depositor-Supplied-Synonyms>

Some of the synonyms of Tylenol contains the phrase “aspirin-free” or “non-aspirin”. Note that Tylenol was returned from a search for “aspirin” (through partial matching using the [synonym] index).

2.3. MeSH Synonyms

The National Library of Medicine (NLM)'s Medical Subject Headings (MeSH)^{8,9} is a controlled vocabulary thesaurus of medical terms arranged in a hierarchical structure. It is used for indexing scientific articles from biomedical journals for PubMed and cataloging medical books, documents, and audiovisual materials, in order to facilitate retrieval of medical information at various levels of specificity.

Many of MeSH terms are chemical names (e.g., for drugs, nutrients, metabolites, toxic chemicals, and so on). PubChem performs an automated annotations of PubChem records with MeSH terms (by means of chemical name matching), creating associations between PubChem records and PubMed articles that share the same MeSH annotation. The MeSH term that match a (depositor-provided) synonym of a compound in PubChem is presented with its entry terms under the “MeSH Synonyms” section of the Compound Summary page of that compound.

Go to the Compound Summary page for CID 171511 via the following link to check the MeSH synonyms and Depositor-supplied Synonyms sections.

<https://pubchem.ncbi.nlm.nih.gov/compound/171511#section=Synonyms>

CID 171511 (magnyl) is a mixture of aspirin and magnesium oxide. Currently (as of February 2017), the “Depositor-Supplied Synonyms” section of this compound does not have any synonym that contains the string “aspirin”. Therefore, CID 171511 is not returned from a search for aspirin with the [completesynonym] or [synonym] index specified. However, because one of its depositor-provided synonyms, “magnyl” matches the MeSH term “magnyl”, PubChem generate MeSH synonyms for this compound, by annotating it with the MeSH term “magnyl” and its entry terms, which can be found via the following link:

<https://meshb.nlm.nih.gov/#/record/ui?ui=C024079>

The resulting MeSH synonyms of CID 171511 includes:

- aspirin, magnesium oxide combination
- 2-(acetyloxy)benzoic acid, magnesium oxide mixture
- Acetard
- Magnyl
- aspirin, magnesium oxide mixture

These MeSH synonyms are listed in the “MeSH Synonyms” section of the Compound Summary page for CID 171511. Note that some of the MeSH synonyms now contain the string “aspirin”. When no Entrez index is specified with the query string, “all” indexed fields are searched. Therefore, the search for “aspirin” *without* the [completesynonym] or [synonym] index does return CID 171511 although no depositor-supplied synonyms contain the string “aspirin”.

3. Additional data retrieval approaches in PubChem

3.1. Classification Browser

The PubChem Classification Browser, which allows the user to navigate or search PubChem records associated to a hierarchical classification system of interest, is available via URL:

<http://pubchem.ncbi.nlm.nih.gov/classification>

The Classification Browser can also be accessed from the [PubChem home page](#) (through the “Services” menu at the top or the “Classification” icon on the right column of the page).

Currently, the Classification Browser can retrieve records annotated with terms in the following classification systems:

- MeSH (Medical Subject Headings)
- ChEBI
- FDA Pharmacological Classification
- KEGG
- LIPID MAPS
- World Health Organization (WHO)’s Anatomical Therapeutic Chemical (ATC)
- World Intellectual Property Organization (WIPO)’s IPC (International Patent Classification)

The Classification Browser provides a powerful way to quickly and visually find a desired subset of PubChem records. The output can be displayed in Tree view or List view.

An important feature of the Classification Browser is that **the Table of Contents presented on the Compound Summary is integrated into the Classification Browser, allowing users to quickly retrieve compounds with a particular type of information available.** For example, the figure below shows how to retrieve all compounds with the boiling point information from PubChem.

Browse PubChem data using a classification of interest, or search for PubChem records annotated with the desired classification/term (e.g., MeSH: phenylpropionates, or Gene Ontology: DNA repair). [More...](#)

The screenshot shows the PubChem Classification Browser interface. At the top, there is a 'Select classification' dropdown menu (1) set to 'PubChem: PubChem Compound TOC'. To its right is a search bar with a 'Keyword' dropdown (2) and a 'Search' button. Below this is a 'Classification description (from PubChem)' section (3) with a detailed text description. Underneath the description are options for 'Data type counts to display' (4) with buttons for 'None', 'Compound', 'Yes', and 'No', and a 'Display zero count nodes?' checkbox. The main area is titled 'Browse PubChem: PubChem Compound TOC Tree' and shows a tree structure with various categories and their counts: PubChem Compound TOC (28,664,775), Agrochemical Information (1,948), Biologic Description (617,271), Biological Test Results (2,395,801), Biomolecular Interactions and Pathways (48,228), Chemical and Physical Properties (392,853), Crystal Structures (59,986), and Experimental Properties (159,437). The 'Experimental Properties' category is expanded to show sub-nodes: Auto-Ignition (1,023), Boiling Point (3,862) (6), and Caco2 Permeability (131). On the right side, a search results panel (5) displays a list of compounds with their chemical structures, names, molecular weights, formulas, IUPAC names, create dates, and CID numbers. A red arrow (6) points from the 'Boiling Point' node in the tree to the search results panel.

In the example above, users need to expand the Table of Contents tree to locate the boiling point node. However, this task may not be easy to some users who do not have prior knowledge about where the node that they want to find is located in the Table of Contents tree system. To assist these users, the Classification Browser supports a keyword search against the node names and descriptions of the classification trees. For example, the example below shows how to retrieve compounds with the CAS Registry number. Note that this task involves a search for the term “CAS”.

Browse PubChem data using a classification of interest, or search for PubChem records annotated with the desired classification/term (e.g., MeSH: phenylpropionates, or Gene Ontology: DNA repair). [More...](#)

The screenshot shows the PubChem Classification Browser interface. Key features are highlighted with numbered callouts:

- 1:** Select classification dropdown menu.
- 2:** Search selected classification by dropdown menu (Keyword, CAS).
- 3:** Classification description (from PubChem) section.
- 4:** Display zero count nodes? checkbox.
- 5:** View type dropdown menu (Tree, List).
- 6:** CAS count in the 'Other Identifiers' section of the search results.

The search results show a tree view of the PubChem Compound TOC. The 'CAS' count is 336,506. A red arrow points from this count to a detailed view of the first result, 'L-Citrulline monohydrate: L-Ornithine, N5-(aminocarbonyl)-, hydrate (1:1): 1821753-49-4'.

The Classification Browser also supports the **PubChem BioAssay Classification Tree**, providing an additional approach to browse, search, and access the BioAssay data. More detailed information on the Classification Browser is available at the URL:

http://pubchem.ncbi.nlm.nih.gov/classification/docs/classification_help.html

3.2. Identifier Exchange Service

The Identifier Exchange Service can be found at the following URL:

<http://pubchem.ncbi.nlm.nih.gov/idexchange>

This service allows the user to convert one type of identifiers for a given set of chemical structures into a different type of identifiers for identical or similar chemical structures. Currently, it supports seven types of identifiers: CID, SID, InChI, InChIKey, SMILES, synonyms, Registry ID. When Registry ID is selected as an input or output identifier type, the DSN (Data Source Name) should also be provided.

The input identifier list may be provided using a string, a text file, or Entrez history. When a service request is submitted, it will be queued on PubChem servers. Once the actual task starts to run, the input identifiers will be converted into CIDs (called input CIDs) during the computation, and the CIDs (called output CIDs) that satisfy the condition specified by one of the following operation types will be retrieved:

- **Same CID:** Same CIDs as input CIDs.
- **Same, Stereochemistry:** CIDs that have same stereo centers as input CIDs.
- **Same, Isotopes:** CIDs that have the same isotopes as input CIDs.
- **Same, Connectivity:** CIDs that have the same connectivity as input CIDs.
- **Same parent:** CIDs that have the same parents as input CIDs.
- **Same parent, Stereochemistry:** CIDs that have the same stereo centers and parents as input CIDs.
- **Same parent, Isotopes:** CIDs that have the same isotopes and parents as input CIDs.
- **Same parent, Connectivity:** CIDs that have the same connectivity and parents as input CIDs.
- **Similar 2D compounds:** CIDs similar to the input CIDs in PubChem's 2-D similarity.
- **Similar 3D conformers:** CIDs similar to the input CIDs in PubChem's 3-D similarity.

These output CIDs are then converted into the identifier type specified by the user and written into a file or sent to **Entrez history**. In practice, the identifier exchange service may be used as a quick approach to search the PubChem Compound database using multiple queries, although this type of task may be performed programmatically (for example, using PUG-REST,¹⁰ which will be discussed in Module 7). A more detailed information is available at the URL:

<http://pubchem.ncbi.nlm.nih.gov//idexchange/idexchange-help.html>

3.3. The PubChem Data Sources page

As discussed in [Module 4](#), the PubChem Data Sources page (<https://pubchem.ncbi.nlm.nih.gov/sources/>) helps users determine who provided what information. This page can be used to retrieve the data provided by a data depositor or to download the annotations collected from a data source. For example, the following figure illustrates how to download the boiling point data collected from DrugBank.¹¹

Data Sources

PUBCHEM > DATA SOURCES

1 sources Download ▾

Filter by Sort by Last Updated Latest First ▾

Data Type	Source	Data Counts by Type	Last Updated
<input type="checkbox"/> Live Substances(1) <input type="checkbox"/> Annotations(1) <input type="checkbox"/> Live BioAssays(0) <input type="checkbox"/> On Hold BioAssays(0) <input type="checkbox"/> Classifications(0) <input type="checkbox"/> On Hold Substances(0)	DrugBank Research and Development, Curation Efforts	7,216 Live Substances 27,565 Annotations	2017/02/01



DrugBank

PUBCHEM > DATA SOURCES > DRUGBANK > ANNOTATIONS

The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information.

34 annotation topics Download ▾

27,565 total annotation data items

Absorption, Distribution and Excretion	Download ▾
Action	Download ▾
Biological Half-Life	Download ▾
Boiling Point	Download ▾ JSON Save Display XML Save Display ASNT Save Display
Caco2 Permeability	
Carrier	
CAS	

To obtain a particular kind of annotated information (e.g., boiling points) through the PubChem Data Sources page, one may need to know “in advance” which depositors provide that information. This can be done through a PUG-REST request¹⁰ (to be discussed in detail in Module 7). For example, the following PUG-REST request returns all data sources that provide the boiling point information for chemicals.

<https://pubchem.ncbi.nlm.nih.gov/rest/pug/annotations/heading/boiling%20point/TXT>

On the other hand, one may want to know what kind of information is provided by a given data source. This can also be done using a PUG-REST request:

<https://pubchem.ncbi.nlm.nih.gov/rest/pug/annotations/sourcename/DrugBank/TXT>

This example retrieves all types of annotations collected from DrugBank.

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.
- (8) Medical Subject Headings (MeSH) (<https://www.nlm.nih.gov/mesh/>) (Accessed on.
- (9) Medical Subject Headings (MeSH®) Fact Sheet (<https://www.nlm.nih.gov/pubs/factsheets/mesh.html>) (Accessed on.
- (10) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Bryant, S. H. *Nucleic Acids Res.* **2015**, *43*, W605.
- (11) Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A. C.; Liu, Y. F.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A.; Gabriel, G.; Ly, C.; Adamjee, S.; Dame, Z. T.; Han, B. S.; Zhou, Y.; Wishart, D. S. *Nucleic Acids Res.* **2014**, *42*, D1091.