



Patent and Intellectual Property Information at NCBI

Searching for nucleotide or protein sequences and compounds from patents

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

How does NCBI get patent information?

Sequences associated with patents are incorporated in GenBank through arrangements with the U.S. Patent and Trademark Office (USPTO), and via our partners in the collaborating International Nucleotide Sequence Database Collaboration (INSDC) databases ([EMBL-EBI](#) and [DDBJ](#)), who contribute information that originates from patent offices other than the USPTO.

USPTO-originating patent sequence records for granted patents are provided to NCBI/GenBank by the USPTO via a special pipeline for inclusion within the [Patent division of GenBank](#). During processing of the data, we filter out records based on the following criteria:

- Nucleotide sequences less than 10 bases and those consisting solely of 'n' residues
- Protein sequences less than 4 residues and those consisting solely of 'X' residues
- Indirect sequences: Patents that have been granted for analysis pipelines/algorithms with greater than 250,000 sequences trigger a manual review at NCBI. If it is determined that the sequence data results from the application of an algorithm to already-available public data, the sequences are not included in GenBank.
- Pre-grant sequences: Although the special pipeline for USPTO patent data includes sequence records for "pre-grant" patents, their sequences are not currently incorporated into GenBank due to a variety of technical constraints.

Note that if data problems are encountered for a given USPTO patent, this may temporarily delay the inclusion of the associated sequences into GenBank's Patent division.

Sequence records for patents originating from the European Patent Office (EPO) and Japan Patent Office (JPO) are submitted to INSDC collaborators EBI and DDBJ, respectively. These records are then provided to NCBI via the daily INSDC data-sharing exchange mechanism, and are included in the Patent division of GenBank. NCBI Website users can find USPTO, EPO and JPO patent sequences in the [Nucleotide](#) or [Protein](#) databases. In addition, the sequences released in the Patent division of GenBank have been used to populate a "Patent database" that is searchable using the [BLAST](#) sequence similarity search tool.

Patent information for **chemicals** is incorporated into the PubChem database of chemicals and biological assays by a variety of mechanisms. Several databases that collect patent information have submitted information on chemical substances and have provided related patent identifiers, including [SureChEMBL](#), [IBM](#), [ScripDB](#), [NextMove Software](#), and [BindingDB](#). In addition, some submitters of chemical substance data have included patent information, which may or may not include identifiers. The PubChem staff also collects and incorporates information about patents from [Google's USPTO Bulk Download](#) and [EPO Espacenet](#) services.

NCBI Website users can find patent information provided by submitters on records in the [PubChem Substance](#) database, or in the [PubChem Compound](#) database on chemical compound summary pages, which aggregate all available information based on a common chemical structure.

Note

Although our aim is to be as informative as possible, we cannot guarantee comprehensiveness of patent-related information on the NCBI website. The information we provide is based on data submitted to us. A search for patent information at NCBI is *not to be considered a definitive search for legal purposes*. Our primary objective is to serve as a source of information for researchers interested in biological sequences and chemicals, not to serve as an authoritative patent database for patent searches or investigations.

The NCBI-related dates on these records have *no relationship* with the first release of the information to the public or with any of the dates associated with the submission or processing of a patent by a patent organization. For specific date-related information, we request that users find the patent number, which is often displayed on our records, and then use this identifier to search the database of the originating patent organization (for example: [USPTO](#), [EPO](#), [JPO](#), [WIPO](#)).

What do dates on the sequence and chemical structure records mean?

For sequences, information submitted to GenBank is displayed as Nucleotide and/or Protein database records. As noted above, this NCBI date does not bear any definitive relationship to the date the information may have first been released to the public or to the date associated with the patent record. The date that the record was first available from the NCBI website is displayed at the top of each record (Figure 1).

Revision History ▾

[Sequence 3 from patent US 8110185](#)

7,185 bp linear DNA <https://go.usa.gov/xnTzP>

Accession: GZ123594.1 GI: 378453721

Current status: live

I	II	Version	GI	Accession	Update Date	Action
●		1	378453721	GZ123594.1	Feb 24, 2012 12:11 PM	

Accession [GZ123594](#) was first seen at NCBI on Feb 24, 2012 12:11 PM

Figure 1. A sequence record displaying data submitted to NCBI/GenBank by the USPTO, with the date on which it was first visible on the NCBI website boxed in red. That date displayed in Revision History format is shown to the left.

NCBI Resources ▾ How To ▾

Nucleotide Nucleotide ▾

Advanced

GenBank ▾

Sequence 3 from patent US 8110185

GenBank: GZ123594.1

[FASTA](#) [Graphics](#)

Go to: ▾

LOCUS GZ123594 7185 bp DNA linear **PAT 24-FEB-2012**

DEFINITION Sequence 3 from patent US 8110185.

ACCESSION GZ123594

VERSION

KEYWORDS <https://www.ncbi.nlm.nih.gov/nuccore/GZ123594.1>

SOURCE

ORGANISM Unknown.

For chemicals, PubChem Substance records display information about chemical substances as provided to NCBI by researchers, universities, companies, etc. (Figure 2A). For each record, a "Deposit Date" is recorded when the submitter has officially provided their information to the PubChem staff. A "Hold Date" may be requested by the submit-

ter, which will prevent the release of the information to the public servers until either the "Hold Date" passes or the submitter requests the release of the information. The "Available Date" indicates the date when the record was first queued up to be moved to the production servers. The record is generally viewable by the public within 24-48 hours after queuing. We do not have information on the exact date and time that a record was publicly viewable. PubChem Substance records can be modified by the original submitters and a "Modify Date" will be included when this updated information is queued up to be moved to the production server. All dates are displayed in the "Identification" section of PubChem Substance records. If the record has been modified since the initial release, earlier versions of the record are viewable by clicking on the "Version" link within the "Modify Date" table. PubChem Compound records are created by the PubChem staff to serve as aggregates of all available information based on a common chemical structure (Figure 2B). A "Create Date" indicates the first time that record corresponding to this chemical structure has been queued up to be moved to the production servers, akin to "Available Date" in PubChem Substance records. The record is generally viewable by the public within 24-48 hours after queuing. We do not have information on the exact date and time that a record was publically available. Please note that the PubChem Compound records are continually updated as new relevant information is submitted or acquired. We do not track updates to PubChem Compound records, but report a "Modify Date" when the entire database has been most recently updated.

NIH NLM National Center for Biotechnology Information

PubChem OPEN CHEMISTRY DATABASE

Search Substances

Substance Record for SID 129205537 **A**

83F8888F1F88128AA0B368FAC06BBA5A <https://pubchem.ncbi.nlm.nih.gov/substance/129205537>

STRUCTURE LITERATURE

PubChem SID: 129205537

PubChem CID: 1983 (acetaminophen) Related Compounds...

External ID: 83F8888F1F88128AA0B368FAC06BBA5A

Source: IBM

Version: 6 ▾ Revision History...

2.5 Modify Date	
Version 1	2011-12-04
Version 2	2012-02-25
Version 3	2012-06-14
Version 4	2015-03-17
Version 5	2015-06-25
Version 6 - currently shown	2017-01-25

Please note that the substance record below is presented as provided to PubChem by the source(depositor). For standardized chemical structure and/or annotation information, please visit the compound summary page for acetaminophen.

PUBCHEM > SUBSTANCE > SID 129205537 Available Date: 2011-12-04; Deposit Date: 2011-12-04

NIH NLM National Center for Biotechnology Information

PubChem OPEN CHEMISTRY DATABASE

Search Compounds

Compound Summary for CID 1983 **B**

Acetaminophen <https://pubchem.ncbi.nlm.nih.gov/compound/1983>

STRUCTURE VENDORS DRUG INFO PHARMACOLOGY LITERATURE PATENTS BIOACTIVITIES

PubChem CID: 1983

Chemical Names: Acetaminophen; 4-Acetamidophenol; Paracetamol; 103-90-2; Tylenol; N-(4-Hydroxyphenyl)acetamide; More...

Molecular Formula: C₉H₉NO₂ or HOCC₆H₄NHCOCH₃

Molecular Weight: 151.165 g/mol

InChI Key: RZVAJINJKPORMJF-UHFFFAOYSA-N

Drug Information: Drug Indication Therapeutic Uses Clinical Trials FDA Orange Book FDA UNII

Safety Summary: Laboratory Chemical Safety Summary (LCSS)

Analgesic antipyretic derivative of acetanilide. Acetaminophen has weak anti-inflammatory properties and is used as a common analgesic, but may cause liver, blood cell, and kidney damage. from MeSH

Acetaminophen is a p-aminophenol derivative with analgesic and antipyretic activities. Although the exact mechanism through which acetaminophen exerts its effects has yet to be fully determined, acetaminophen may inhibit the nitric oxide (NO) pathway mediated by a variety of neurotransmitter receptors including N-methyl-D-aspartate (NMDA) and substance P, resulting in elevation of the pain threshold. The antipyretic activity may result from inhibition of prostaglandin synthesis and release in the central nervous system (CNS) and prostaglandin-mediated effects on the heat-regulating center in the anterior hypothalamus.

Acetaminophen is a widely used nonprescription analgesic and antipyretic medication for mild-to-moderate pain and fever. Harmless at low doses, acetaminophen has direct hepatotoxic potential when taken as an overdose and can cause acute liver injury and death from acute liver failure. Even in therapeutic doses, acetaminophen can cause transient serum aminotransferase elevations. from LiverTox Summary from LiverTox

Modify Date: 2017-10-28; Create Date: 2004-09-16

PUBCHEM > COMPOUND > ACETAMINOPHEN Modify Date: 2017-10-28; Create Date: 2004-09-16

Figure 2. A) This is the top section of a PubChem Substance record displaying the number of record updates (versions) and the corresponding date section of the record. B) This is the top section of a PubChem Compound record with the "Create" and "Modify" dates highlighted.

How can records with patent information be found?

Searching for sequences: To access information for nucleotide sequence records included in GenBank's Patent division or their corresponding protein sequences, search the Nucleotide or Protein databases with the terms [gbdiv](#) [pat](#) [\[Properties\]](#) or [patent\[Properties\]](#). To search for sequence records associated with a specific patent, search the Nucleotide or Protein database with (country code|patent number), for example: ["us|8669056"](#).

The image shows two side-by-side screenshots of the NCBI search interface. The left screenshot, labeled 'A', shows a search in the Protein database for 'patent[properties]'. The search results show 509,195 items, with the first four results being protein sequences from patent US 9607202. The right screenshot, labeled 'B', shows a search in the Nucleotide database for 'us|8669056'. The search results show 64,922 items, with the first four results being DNA sequences from patent US 8669056. Both screenshots include a search bar, filters, and a list of results with links to the full records.

Figure 3. A) This is a search in the Protein database for all patent entries. **B)** This is a search in the Nucleotide database for entries from a specific US Patent 8669056.

Searching for chemicals: To access all chemical records with patent information search the PubChem Compound database with [has_patent\[Filter\]](#). To search for chemical records associated with a specific patent, search the PubChem Compound database with the quoted patent id in this format ["us9481677"](#). Note that there is no pipe symbol ("|") separating the country and the patent id. These strategies also apply to the PubChem Substance database.

The image shows two side-by-side screenshots of the NCBI search interface. The left screenshot, labeled 'A', shows a search in the PubChem Substance database for 'has_patent[filter]'. The search results show 326,779,31 items, with the first three results being chemical structures from patent US9481677. The right screenshot, labeled 'B', shows a search in the PubChem Compound database for 'us9481677'. The search results show 2 items, with the first two results being chemical structures from patent US9481677. Both screenshots include a search bar, filters, and a list of results with links to the full records.

Figure 4. A) This is a search in the PubChem Substance database for all entries with patent information. **B)** This is a search in the PubChem Compound database for entries with information from a specific US Patent 9481677.

To search for chemical records associated with a specific patent, e.g., EP0521471, you can also go to the Patent View page for that patent via the URL containing the patent number at the end. The URL for the above patent is <https://pubchem.ncbi.nlm.nih.gov/patent/EP0521471>.

The Patent View page may also be reached through the PubChem Search page (<https://pubchem.ncbi.nlm.nih.gov/search>). It allows one to search for patents by patent number, inventor, applicant, and chemical name/structure/pattern.

How can records with patent information be found? (cont.)

Searching with sequences: the [BLAST sequence search tool](#) was designed to search for identical or similar [nucleotide](#) or [protein](#) sequences from those in a specified database. A "Patent database" is available containing all

the sequences in Genbank's Patent division as mentioned earlier. Simply type or paste in a sequence or accession into the query box and select this database in the pull-down menu to perform your search. To determine when this database was most recently updated and the total number of sequences within, click the blue question mark button next to the database pull-down menu (Figure 5). For more information, please see the comprehensive set of [Help documents and videos](#).

The screenshot shows the BLAST Standard Nucleotide BLAST interface. The 'Choose Search Set' section is highlighted with a red box and a yellow circle 'A'. The 'Database' dropdown is set to 'Patent sequences(pat)'. The 'Program Selection' section shows 'Optimize for' set to 'Highly similar sequences (megablast)'. The 'BLAST' button is visible at the bottom.

Figure 5. A) A Nucleotide BLAST form with database set to "Patent sequences (pat)." The short URL links to a live page with the same setting. **B)** The "Choose Search Set" section of a Protein BLAST form, with database set to "Patented protein sequences (pat)." The short URL links to a live page with the same setting.

The screenshot shows the 'Choose Search Set' section for a Protein BLAST form. The 'Database' dropdown is set to 'Patented protein sequences(pat)'. The 'Title' is 'Protein sequences derived from the Patent division of GenBank'. The 'Molecule Type' is 'Protein'. The 'Update date' is '2017/11/02'. The 'Number of sequences' is '2113860'. A short URL 'https://go.usa.gov/xnTXP' is provided.

Searching with chemical structures: the [PubChem Structure Search](#)

[tool](#) was designed to search for identical or similar chemical structures from those in the PubChem Compound database. A large variety of methods for inputting your query are available, including names, SMILES, InChIs, Chemical structure files, and you can even draw in a structure using the "PubChem Structure Editor/Sketcher". Resulting identical or similar structures will be retrieved in the PubChem Compound interface for further discovery. For more information, please see the extensive [PubChem Structure Search Help documentation](#). To filter the results of a PubChem Structure Search to those records which contain patent information, you can use the "Search History" to combine the results of a structure search with the results of a filtering of PubChem Compound records containing patent information (as mentioned above).

To do this, run the structure search. Then click on "Advanced" below the search text box at the top of the page. In the "Search Builder" section, click a pull-down menu and select "Filter" and type "has patent" into the corresponding text box. Then scroll down to see the History table and "Add" the previously run search to the builder (it will look like "Select XXXX document(s)", with the X's representing the number of structures retrieved in the structure search). Then, press the "Search button". For more information on using the "Advanced Search" page, please see the [Advanced Search Help documentation and video](#)

The screenshot shows the PubChem Structure Search interface. The 'Search By' dropdown is set to 'Name/Text'. The 'Search' button is visible. A short URL 'https://pubchem.ncbi.nlm.nih.gov/search/search.cgi#' is provided.

Figure 6. PubChem structure search interface. Use different tabs for specific type of input. A structure sketcher/drawing tool is available under the "Identity/Similarity," "Substructure/Superstructure," and "3D Conformer" tabs. The direct link to this utility is <https://pubchem.ncbi.nlm.nih.gov/edit2/index.html?cnt=0>