

PubChem

PubChem is a database of chemical molecules and their activities against biological assays. The system is maintained by the National Center for Biotechnology Information (NCBI), a component of the National Library of Medicine, which is part of the United States National Institutes of Health (NIH). PubChem can be accessed for free through a web user interface. Millions of compound structures and descriptive datasets can be freely downloaded via FTP. PubChem contains multiple substance descriptions and small molecules with fewer than 100 atoms and 1000 bonds. More than 80 database vendors contribute to the growing PubChem database.^[2]

Contents

Databases

Searching

History

ACS's concerns

Database fields

See also

References

External links

PubChem



Content

Description Chemicals and their bioassays

Data types captured dhsh

Organisms Humans and other animals

Contact

Research center NCBI

Primary citation PMID 15879180 (https://pubmed.ncbi.nlm.nih.gov/15879180)

Access

Website https://pubchem.ncbi.nlm.nih.gov/

Download URL FTP (https://pubchem.ncbi.nlm.nih.gov/pc_fetch/pc_fetch-help.html)

Web service URL PUG-View (https://pubchemdocs.ncbi.nlm.nih.gov/pug-view)^[1]

Miscellaneous

License Public domain

Databases

PubChem consists of three dynamically growing primary databases. As of 5 November 2020 (number of BioAssays is unchanged):

- Compounds, 111 million entries^[3] (up from 94 million entries in 2017^[4]), contains pure and characterized chemical compounds.^[5]
- Substances, 293 million entries^[3] (up from 236 in 2017^[6] and 163 million entries in Sept 2014^[7]), contains also mixtures, extracts, complexes and uncharacterized substances.
- BioAssay, bioactivity results from 1.25 million^[8] (up from 6000 in Sept 2014^[9]) high-throughput screening programs with several million values.

Searching

Searching the databases is possible for a broad range of properties including chemical structure, name fragments, chemical formula, molecular weight, XLogP, and hydrogen bond donor and acceptor count.

PubChem contains its own online molecule editor with SMILES/SMARTS and InChI support that allows the import and export of all common chemical file formats to search for structures and fragments.

Each hit provides information about synonyms, chemical properties, chemical structure including SMILES and InChI strings, bioactivity, and links to structurally related compounds and other NCBI databases like PubMed.

In the text search form the database fields can be searched by adding the field name in square brackets to the search term. A numeric range is represented by two numbers separated by a colon. The search terms and field names are case-insensitive. Parentheses and the logical operators AND, OR, and NOT can be used. AND is assumed if no operator is used.

Example (Lipinski's Rule of Five):

```
0:500[mw] 0:5[hbdc] 0:10[hbac] -5:5[logp]
```

History

PubChem was released in 2004.^[10]

ACS's concerns

The American Chemical Society has raised concerns about the publicly supported PubChem database, since it appears to directly compete with their existing Chemical Abstracts Service.^[11] They have a strong interest in the issue since the Chemical Abstracts Service generates a large percentage of the society's revenue. To advocate their position against the PubChem database, ACS has actively lobbied the US Congress.

Soon after PubChem's creation, the American Chemical Society lobbied U.S. Congress to restrict the operation of PubChem, which they asserted competes with their Chemical Abstracts Service.^[12]

Database fields

Identification numbers

- Identification number in current database [UID]
- Substance identification number [SID]
- Compound identification number [CID]
- BioAssay identification number [BAID], [AID]

General

- Any database field [ALL]
- Comment [CMT]
- Deposition date [DDAT], [DEPDAT]
- Depositor's external ID [SRID], [SRCID]
- Source name [SRC], [SRCNAM], [SRCNAME]
- Source release date [SRD], [SRDAT], [RLSDAT]
- Medical Subject Heading (MeSH) term [MSHT], [MESHT]

- MeSH tree node [MSHN], [MESHTN]
- MeSH pharmacological actions [PHMA], [PHARMA]

Substance properties

- Substance synonyms [SYNO]
- IUPAC name [UPAC], [IUPAC]
- International Chemical Identifier (InChI) [INCHI]
- Molecular weight [MW], [MWT], [MOLWT]
- Chemical elements [ELMT], [EL]
- Non-Hydrogen atoms [HAC], [HACNT]
- Isotope count [IAC], [IACNT]
- Total formal charge [TFC], [CHG], [CHRG]
- Chiral atom count [ACC], [ACCNT]
- Defined chiral atom count [ACDC], [ACDCNT]
- Undefined chiral atom count [ACUC], [ACUCNT]
- Hydrogen bond acceptor count [HBAC], [HBACNT]
- Hydrogen bond donor count [HBDC], [HBDCNT]
- Tautomer count [TC], [TCNT], [TTMC]
- Rotatable bond count [RBC], [RBCNT]
- XLogP^[13] [XLGP], [LOGP]

Compound properties

- Compound synonyms [CSYN], [CSYNO]
- Component count [CC], [CCNT]
- Covalent unit (molecule) count [CUC], [CUCNT]
- Total bioactivity count [TAC]

See also

- Chemical database
 - CAS Common Chemistry - run by the American Chemical Society
 - Comparative Toxicogenomics Database - run by North Carolina State University
 - ChEMBL - run by European Bioinformatics Institute
 - ChemSpider - run by UK's Royal Society of Chemistry
 - DrugBank - run by the University of Alberta
 - IUPAC - run by Swiss-based International Union of Pure and Applied Chemistry (IUPAC)
 - Moltable - run by India's National Chemical Laboratory
 - PubChem - run by the National Institute of Health, USA
 - BindingDB - run by the University of California, San Diego
 - SCRIPDB - run by the University of Toronto, Canada
 - National Center for Biotechnology Information (NCBI) - run by the National Institute of Health, USA
 - Entrez - run by the National Institute of Health, USA
 - GenBank - run by the National Institute of Health, USA

References

1. Kim, Sunghwan; Thiessen, Paul A.; Cheng, Tiejun; Zhang, Jian; Gindulyte, Asta; Bolton, Evan E. (9 August 2019). "PUG-View: programmatic access to chemical annotations integrated in PubChem" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6688265>). *Journal of Cheminformatics*. **11** (1): 56. doi:10.1186/s13321-019-0375-2 (<https://doi.org/10.1186%2Fs13321-019-0375-2>). PMC 6688265 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6688265>). PMID 31399858 (<https://pubmed.ncbi.nlm.nih.gov/31399858>).
2. "PubChem Source Information" (<https://pubchem.ncbi.nlm.nih.gov/sources/sources.cgi>). *The PubChem Project*. USA: National Center for Biotechnology Information.
3. Kim, Sunghwan; Chen, Jie; Cheng, Tiejun; Gindulyte, Asta; He, Jia; He, Siqian; Li, Qingliang; Shoemaker, Benjamin A; Thiessen, Paul A; Yu, Bo; Zaslavsky, Leonid; Zhang, Jian; Bolton, Evan E (8 January 2021). "PubChem in 2021: new data content and improved web interfaces" (<https://doi.org/10.1093%2Fnar%2Fgkaa971>). *Nucleic Acids Research*. **49** (D1): D1388–D1395. doi:10.1093/nar/gkaa971 (<https://doi.org/10.1093%2Fnar%2Fgkaa971>).
4. "Search Results for all compounds" (<https://www.ncbi.nlm.nih.gov/pccompound?term=all%5Bfilt%5D&cmd=search>). Retrieved 28 January 2016.
5. "all[**filt**] - PubChem Compound Results" (<https://www.ncbi.nlm.nih.gov/sites/entrez?term=all%5Bfilt%5D&cmd=search&db=pccompound>). *The PubChem Project*. USA: National Center for Biotechnology Information. Retrieved 7 January 2011.
6. "all[**filt**] - PubChem Substance Results" (<https://www.ncbi.nlm.nih.gov/sites/entrez?term=all%5Bfilt%5D&cmd=search&db=pcsubstance>). *The PubChem Project*. USA: National Center for Biotechnology Information. Retrieved 28 January 2016.
7. "all[**filt**] - PubChem Substance Results" (<https://www.ncbi.nlm.nih.gov/sites/entrez?term=all%5Bfilt%5D&cmd=search&db=pcsubstance>). *The PubChem Project*. USA: National Center for Biotechnology Information. Retrieved 7 January 2011.
8. "all[**filt**] - PubChem BioAssay Results" (<https://www.ncbi.nlm.nih.gov/sites/entrez?term=all%5Bfilt%5D&cmd=search&db=pcassay>). *The PubChem Project*. USA: National Center for Biotechnology Information. Retrieved 28 January 2016.
9. "all[**filt**] - PubChem BioAssay Results" (<https://www.ncbi.nlm.nih.gov/sites/entrez?term=all%5Bfilt%5D&cmd=search&db=pcassay>). *The PubChem Project*. USA: National Center for Biotechnology Information. Retrieved 7 January 2011.
10. "About PubChem" (<https://pubchemdocs.ncbi.nlm.nih.gov/about>). Retrieved 3 May 2014.
11. Kaiser J (May 2005). "Science resources. Chemists want NIH to curtail database". *Science*. **308** (5723): 774. doi:10.1126/science.308.5723.774a (<https://doi.org/10.1126%2Fscience.308.5723.774a>). PMID 15879180 (<https://pubmed.ncbi.nlm.nih.gov/15879180>). S2CID 166918466 (<https://api.semanticscholar.org/CorpusID:166918466>).
12. "PubChem and the American Chemical Society" (<https://osc.universityofcalifornia.edu/2005/05/american-chemical-society-calls-on-congress-to-shut-down-nih-pubchem/>). *Reshaping Scholarly Communication*. USA: University of California. 2005-05-31. Retrieved 2018-10-15.
13. Cheng T (Nov 2007). "Computation of octanol-water partition coefficients by guiding an additive model with knowledge". *Journal of Chemical Information and Modeling*. **47** (6): 2140–2148. doi:10.1021/ci700257y (<https://doi.org/10.1021%2Fci700257y>). PMID 17985865 (<https://pubmed.ncbi.nlm.nih.gov/17985865>).

External links

- Official website (<https://pubchem.ncbi.nlm.nih.gov>)

This page was last edited on 24 January 2021, at 18:28 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.