

Volně dostupné chemické databáze

(Rich Apodaca, Thirty-Two Free Chemistry Databases – <http://depth-first.com/articles/2007/01/24/thirty-two-free-chemistry-databases> - doplněno)

PubChem

PubChem- The granddaddy of all free chemistry databases. Search over 8 million compounds by a variety of criteria. Although some PubChem records are linked into the primary literature through MeSH, most are not. But this doesn't seem to be PubChem's true calling. Instead, PubChem may well evolve into the world's largest online collection of molecular data sheets. Increasingly, the other databases in this list are cross-referencing their entries into PubChem. PubChem's entire database can be [downloaded by FTP](#). [CAS Registry](#) are correct to see PubChem as the first real competition they've had in decades.



ZINC- A free database of commercially-available compounds for virtual screening. Search over 4.6 million compounds by structure, IUPAC name, InChI, and a host of calculated properties. For noncommercial purposes, the ZINC database may be downloaded in whole or in part for local use.



eMolecules- Google for molecules. With a simple interface and super fast search engine, eMolecules augments PubChem with other information sources, including specialty chemical catalogs. Although eMolecules' emphasis seems to be on commercially-available compounds, it's only possible to get a link directly into a supplier's online catalog for a limited number of molecules. Most of the links are to PubChem records. For this reason, I don't find eMolecules very useful in its current form. If you remember something called "Choogle", this is the same service (moral: [don't mess with Google](#)).



CHEBI- "A freely available dictionary of molecular entities focused on 'small' chemical compounds." CHEBI draws its information from [two main sources](#): Integrated Relational Enzyme Database of the EBI and the Kyoto Encyclopedia of Genes and Genomes. Find out what proteins a molecule has been associated with and in what context. Provides cross-links to CAS registry numbers, Beilstein registry numbers, and Gmelin registry numbers.

NIST

NIST Chemistry WebBook- Physical data (thermochemical, thermophysical, and ion energetics) for mostly organic compounds. Search by formula, structure, CAS number, and IUPAC name.



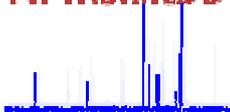
BioCyc- A collection of about 3,500 compounds involved as enzyme substrates, products, inhibitors, and activators. On accepting a license agreement, the entire database can be [freely downloaded](#) in [Chemical Markup Language](#) format.



ChemExper- Find a supplier for your specialty chemical needs. Search by structure, name, molecular formula, and CAS number. After finding your compound, get an offer from one or more suppliers. I can't vouch for how this works in practice, but it sounds like a good idea.

Compendium of Pesticide Common Names- More than 1,100 commonly-used pesticides. Compounds are located by browsing indexed lists (IUPAC name, CAS number, and trade name) rather than searching. Each entry lists, among other pieces of information, a chemical structure and sub-classifications (repellents, antifeedants, synergists, etc.).

NMRShiftDB



NMRShiftDB- Organic structures and their nuclear magnetic resonance (nmr) chemical shifts. NMRShiftDB contains chemical shift data for over 22,000 organic compounds and 19,000 spectra. Records can be searched by structure, chemical shift and nucleus. NMRShiftDB is truly open; it can be [accessed programmatically](#) and the source code for the software that runs the online database can be [freely downloaded](#). Individual users can submit their own spectral shifts for peer review and subsequent inclusion.

Chemical Structure Lookup Service (CSLS)- An address book for chemical structures. If you've ever used [Metacrawler](#), then you'll recognize the idea behind CSLS, which is to aggregate several free chemistry databases. Search over 27 million molecules by IUPAC name, InChI, structure, SMILES, and a variety of molecular identifiers. Your results set will contain links into specific databases that host the molecules you find. The user interface isn't just unfriendly - it's downright antisocial. But if you can get past this, CSLS may well be one of the most useful services in this list.

DrugBank

DrugBank- Combines detailed drug data with comprehensive drug target information. Search over 4,300 drugs by trade name, SMILES, and InChI. Each record contains information on target of action, therapeutic indication, medications the drug is an ingredient of, and trade names. Searches can be limited to only approved drugs or experimental drugs. Both the concept and interface to this service are well thought-out. *Note: this service was unavailable as of Jan 19, 2007*



Wikipedia- Wikipedia? Yes, Wikipedia. Wikipedia offers several kinds of chemical information produced by a knowledgeable, all-volunteer army. Looking for information on organic compounds? Consider [this datasheet on morphine](#) as an example. For those interested in synthesis, Wikipedia is increasingly being used to collaboratively author [short reviews on the topic](#). Search capabilities are currently limited to text and don't appear to work with IUPAC names or CAS numbers. Where this quintessential [disruptive technology](#) and its offspring end up taking chemical publishing is unclear, but the ride will be spectacular. Newly [Wikipedia Chemistry Portal](#) was established.



Wikipedia:WikiProject Chemicals/Data is a collection of links and references that are useful for chemistry-related works. WikiProject Chemicals/Data includes free online [chemical databases](#), publications, [patents](#), [computer programs](#), and various tools.

ChemDB- A chemical database is but one of the services offered by this site. Search over 4.1 million compounds by structure, or various calculated properties. ChemDB also offers a variety of free online cheminformatics tools such as Babel file format conversion, SMILES depict, and molecular property calculation. Read more about ChemDB in [this Bioinformatics paper](#).

CHEMBANK

ChemBank- Structure search over 36,000 original biological assays of small molecules collected by Harvard's [Institute of Chemistry and Cell Biology](#) (ICCB). Many of the data contained in ChemBank have never been published, making this database particularly valuable.



National Institute of Allergy and Infectious Diseases Database- Structure search hundreds of thousands of screening datapoints collected by the NIAID in its HIV, Opportunistic Infection, and TB programs.



[National Toxicology Program](#)- Search by name for compounds listed in the NTP database. Returns detailed internal reports and links to the primary literature.



[NIST Chemical Kinetics Database](#)- Search by reagent or product name or formula for gas phase rate constants collected from the primary literature.



[Computational Chemistry Comparison and Benchmark Database](#)- Search by formula for over 600 gas phase atom and molecule physical chemistry data obtained experimentally and by computation.

NIST

[IUPAC-NIST Solubility Data Series](#)- Search by name or CAS number through over 67,000 solubility measurements. Data were comprehensively compiled from over 1,800 references in primary literature.

SOLV-DB[®]

[SOLV-DB](#)- Search over 200 common solvents by name, CAS number, or chemical formula physical. Available data include boiling point, water solubility, viscosity, octanol-water partition constant, flash point, and a variety of other properties.

PDSP K_i Database

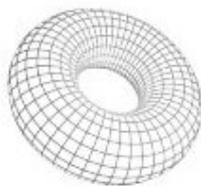
[NIMH Psychoactive Drug Screening Program K_i Database](#)- Search over 44,000 K_i determinations culled from the literature. Although this database appears to have no structure search capability, this is listed as a "Future Enhancement". This is a perfect example of a very useful service that could do with a major user interface redesign. There also appears to be another (defunct) service by the same name, but a [different URL](#).



[Kyoto Encyclopedia of Genes and Genomes \(KEGG\)](#)- A Japanese counterpart to PubChem/PubMed. One of the most interesting services on this list, KEGG consists of four interconnected databases: KEGG Pathway; KEGG genes; KEGG Brite; and KEGG Ligand. KEGG Ligand contains over 14,000 compounds searchable by name, and crosslinked to over 45,000 biological pathways. The KEGG Ligand database can be searched by structure through [KegDraw](#), a 2-D structure editor written in Java. With some minor configuration on my Linux system, I was able to perform some basic substructure searches using KegDraw. Your mileage may vary. A nice overview of KEGG is available in a [recent article](#). The contents of KEGG can be [downloaded by anonymous ftp](#) for academic use.

BRENDA

[BRENDA](#)- Search over 40,000 structures as substrates, products, cofactors, or inhibitors for enzymes. Although my search was able to find compounds by substructure, I was not able to view any links to the results. Your mileage may vary.



[Biochemical Pathways Database](#)- Structure search over 1,100 small molecules as participants in biochemical pathways. A potentially useful service, but currently too slow to fully evaluate. A structure search for naphthalene hung for five minutes before I terminated it without success.



[ChemMine](#)- Search by structure for compounds collected from a variety of open databases. View assay results in annotated biological experiments. I find the layout and organization of this service annoyingly confusing, but the underlying information appears to be useful nevertheless. Behind the scenes, ChemMine uses two open source cheminformatics libraries: [Open Babel](#) and [JOELib](#). For a more detailed view of ChemMine, see the [recent article](#).



[Organic Syntheses](#)- Search by structure through the entire contents of synthetic organic chemistry's flagship resource. Substructure search requires Chime, so if you run Linux, or for some other reason can't install the plugin, you'll be out of luck.



[WebReactions](#)- Structure search organic reactions in four databases containing a total of over 391,000 reactions. Each reaction hit is linked to the primary literature through a bibliographical reference. Although the interface takes some getting used to, WebReactions may make a worthy companion to the traditional SciFinder search.



[Spectral Database for Organic Compounds \(SDBS\)](#)- Search by name, molecular formula, molecular weight range, or CAS number through over 14,000 full ^1H NMR spectra, 12,000 full ^{13}C spectra, and 50,000 full FT-IR spectra collected from over 32,000 compounds.



[BindingDB](#)- Structure search over 24,000 K_i and IC_{50} measurements from over 10,000 molecules. Data is collected from, and cross-referenced to, the primary literature. I was unable to determine how to submit a substructure search through the Marvin applet on my Linux system (there is no "Search" button, for example). A text search for "naphthalene", for example, showed some impressive potential for this database. Anyone can currently contribute to BindingDB, one of the few databases on this list to have such a policy.



[PDBbind](#)- Browse over 2,700 complexes of small molecules ligands with proteins found in the Protein Databank. Structure searching requires a license. 3-D rendering comes courtesy of the ever-popular [Jmol](#) applet.



[AffinDB](#)- Search affinity data for complexes found in the Protein Databank. Affinity data are cross-linked to the primary literature through PubMed. Small molecule searching is limited to IUPAC names provided in a

pull-down menu. By registering, users can upload affinity data themselves. AffinDB is just one example of what might be possible as chemistry databases begin to combine multiple sources of data into easy-to-use packages.



[ChemRefer](#)- It doesn't get any simpler. Type in your keywords and get links to the matching full-text PDFs from the primary literature. As [mentioned before](#), the legality of some of ChemRefer's holdings, for example its articles from ACS journals, is not clear. But as more chemistry journals go [Open Access](#), look to services like ChemRefer to play an increasing role in the way scientists navigate the primary literature.



[ChemSpider](#) makes available a database of chemical structures and predicted properties as well as providing access to a series of property prediction algorithms. ChemSpider is built with the intention of aggregating and indexing chemical structures and their associated information into a single searchable repository and making it available to everybody, for free. ChemSpider is fast becoming a community for chemists to share data and knowledge associated with chemical structures and reactions and providing an open forum for discussion and sharing.



[Patent Alert](#) is a free e-mail publication for people interested in patents! Upon subscribing, you will receive periodic updates about inventions **recently patented in the United States**. Each entry will link to the **full-text patent document** at the United States Patent Office website. You may choose from a host of patent update topics such as **Aeronautics, Artificial Intelligence, Automobile, Internet, New Drugs**, and more! Our updates are free, and they do not contain advertising or sponsored content! You can choose to receive updates weekly, bi-weekly or monthly.