*Future Internet* 2010, *2*, 635-644; doi:10.3390/fi2040635

OPEN ACCESS

*future internet*

**ISSN 1999-5903**

www.mdpi.com/journal/futureinternet

*Article*

# CWM Global Search—The Internet Search Engine for Chemists and Biologists

**Alexander Kos * and Hans-Jürgen Himmler**

AKos Consulting & Solutions Deutschland GmbH (AKos GmbH), Austr. 26, D-79585 Steinen, Germany; E-Mail: contact@hjhimmler.de

* Author to whom correspondence should be addressed; E-Mail: software@akosgmbh.de; Tel.: +49-7627-970068; Fax: +49-7627-970067.

*Received: 11 October 2010; in revised form: 25 October 2010 / Accepted: 30 November 2010 / Published: 3 December 2010*

---

**Abstract:** CWM Global Search is a meta-search engine allowing chemists and biologists to search the major chemical and biological databases on the Internet, by structure, synonyms, CAS Registry Numbers and free text. A **meta-search engine** is a search tool that sends user requests to several other search engines and/or databases and aggregates the results into a single list or displays them according to their source [1]. CWM Global Search is a web application that has many of the characteristics of desktop applications (also known as Rich Internet Application, RIA), and it runs on both Windows and Macintosh platforms. The application is one of the first RIA for scientists. The application can be started using the URL http://cwmglobalsearch.com/gsweb.

**Keywords:** internet search engine; meta-search engine; rich internet application; RIA; CWM global search; chemical structure search

---

## 1. Introduction

The Internet is changing how we work [2]. The Internet will also become the major information tool for chemists and biologists. Google has changed how we search and what we expect for answers. Our brain adjusts to the way we work [3]. We have to realize that the Internet is not a simple tool, but something that can control us, and this can be a frightening thought. We get used to always finding an answer, but we need to watch out that we do not forget to search not for the first best, but for the best answer; otherwise we lose on efficiency what we gain on speed.

Chemists have a powerful language—the chemical structure—and this is not in the realm of Google and other search engines such as Wikipedia. Databases cannot be indexed like simple html pages and Google and other search engines cover only a fraction of the knowledge. Databases on the Internet, however, can be searched by web services and CGI (Common Gateway Interface) scripts.

## 2. Product Description

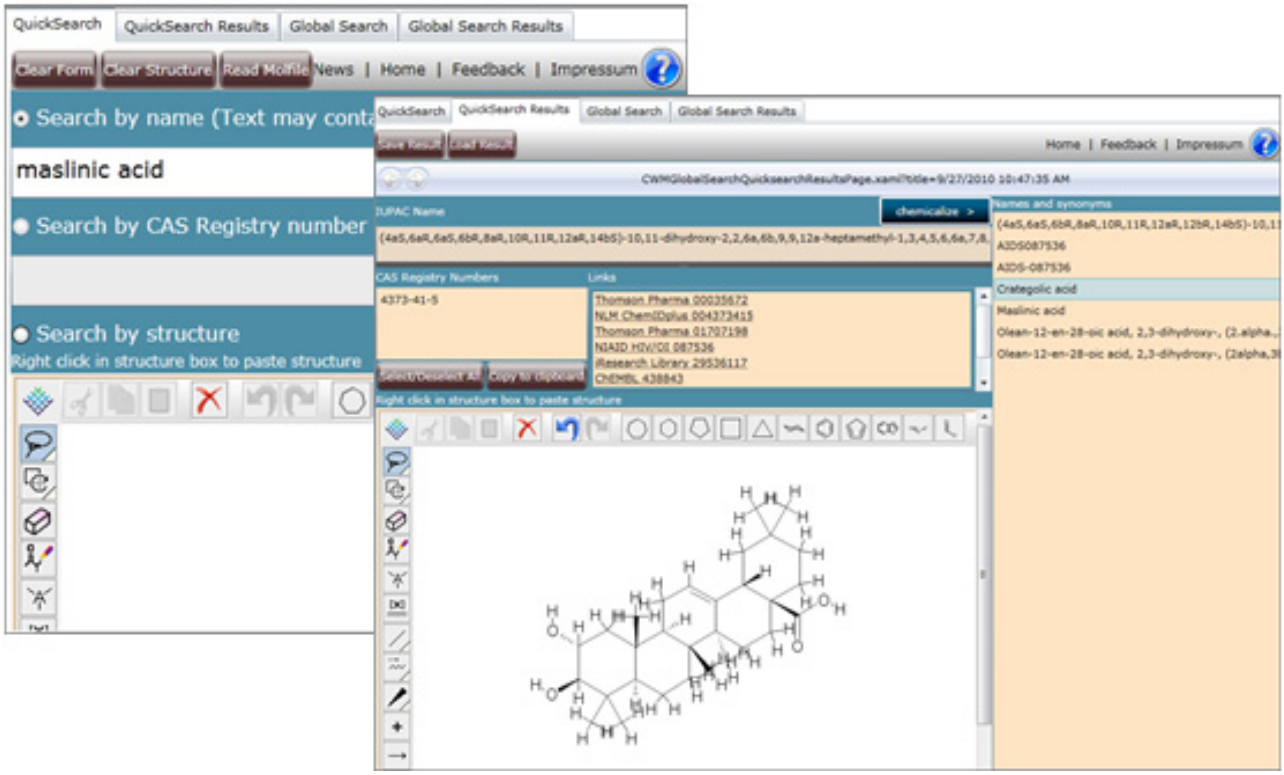
The web application CWM Global Search uses the chemical structure and web services to search Internet websites and Internet databases. CWM Global Search is not a database, and does not rely on an index, but relies on freely accessible web services and freely accessible scripting interfaces to websites [4]. There is no database to update. CWM Global Search will always produce results from the current view of what is available on the Internet.

CWM Global Search allows all major, freely accessible, chemical databases hosted on the Internet to be searched using a chemical structure, CAS Registry number or synonym as a query.

The program generates InChI identifiers (both InChI Names and InChI Keys), molfiles, as well as Smiles strings from the query structure and subsequently locates those identifiers on the Internet.

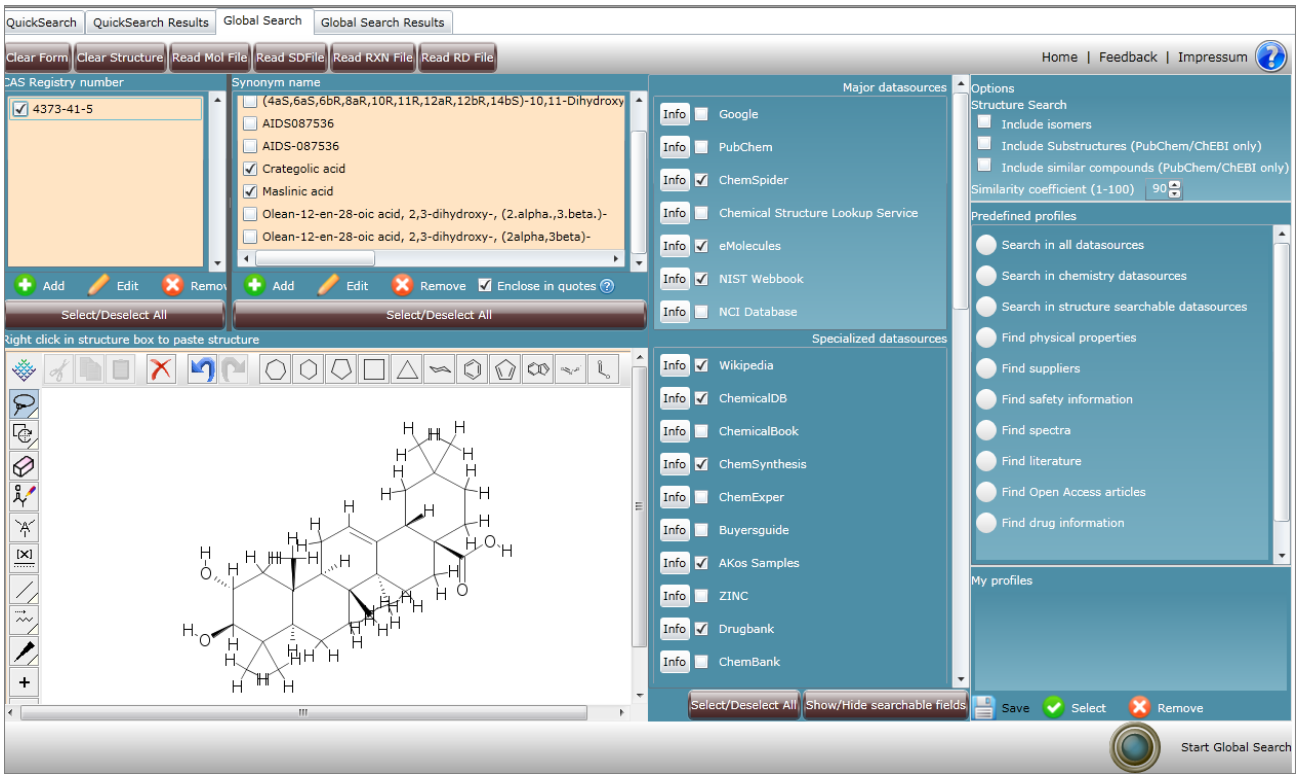
The user interface has a Quick Search and a Global Search. In the Quick Search the user enters structure, name or CAS Registry Number. The program locates the IUPAC name, synonym names, brand names, other known identifiers and CAS Registry Numbers besides the structure using the NCI/CADD Chemical Identifier Resolver [5]. Some hyperlinks are presented that might be already satisfactory for the user. This first search is very fast. An example of the quick search results is shown in Figure 1.

**Figure 1.** The CWM Global Search user interface Quick Search—synonym search for maslinic acid.



The results of the Quick Search, structure, names and CAS Registry Numbers can be selected for a more comprehensive search using the Global Search. The results allow "cross-linking", for example: Quick Search can be used to get a list of synonyms for 'maslinic acid', and then this list can be used in a Global Search. You will find the natural occurrence of this compound in an article in Wikipedia which only mentions 'crategolic acid', (only one of the known synonyms for maslinic acid returned by the QuickSearch), but nowhere does the article in Wikipedia mention 'maslinic acid'.

**Figure 2.** The CWM Global Search user interface Global Search—Combined CAS registry number synonym, and structure search for maslinic acid.




In Global Search, one can search presently over 46 different databases and Google. This includes chemistry centric databases such as ChemSpider, databases specialized in finding commercial suppliers such as eMolecules, databases with a focus on biological data such as PubChem, KEGG, ChEBI and the DrugBank, databases containing patent information such as SureChem, and literature databases such as MedLine. Some sources are gateways, like Open J-Gate, which is an electronic gateway to global journal literature in the open access domain. A complete list of the sources can be found in Appendix 1.

The user can select the data sources using predefined profiles, *i.e.*, chemistry, biology, availability, safety, *etc*., and can create his own profiles, see Figure 2.

The search returns a collection of hyperlinks which allow direct drill down to the corresponding database to look at the data associated with the query. These drill-down pages are parsed by CWM Global Search to generate facts. These are true/false indicators that tell the user what kind of data he potentially can find in the corresponding database for his query. The result of the search is a grid with links and highlighted facts; see Figure 3. Examples of such facts are the presence of commercial

suppliers, availability of safety information or spectra, as well as the presence of known biological activities associated with the query.

**Figure 3.** Global Search Result Page of the combined CAS registry number synonym, and structure search for maslinic acid.

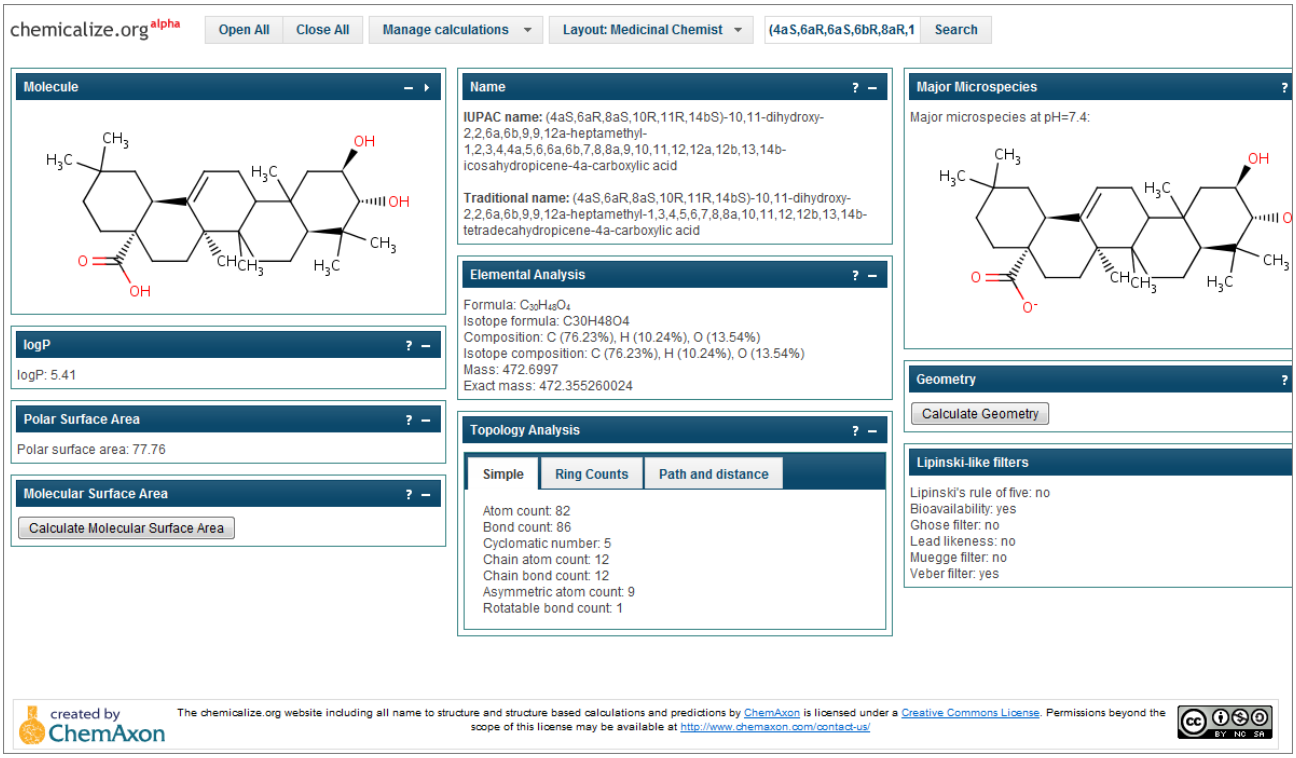

CWM Global Search supports exact structure searches, isomer searches (tautomers and stereoisomers), as well as substructure and similarity searches. At the moment, only PubChem and ChEBI support a substructure and similarity search using the drawn structure without query features. The structures can be copy and pasted from any of the major chemical drawing programs, or can be drawn directly in the JDraw [6] applet.

CWM Global Search supports single structure searches and multiple structure queries via support of SDFiles. In addition, the program supports reaction based queries by support of RXN structures and/or RDFiles. In case a reaction is used as a query, CWM Global Search searches the Internet for all reactants and products contained in the reaction. This is not a reaction search, but is useful for finding suppliers for starting materials, while at the same time also making sure that the product cannot be bought.

## 3. Additional Features

In the Quick Search and Global Search results page, a button “chemicalize” links to a page that displays calculated values for the compound; see Figure 4. Chemicalize (www.chemicalize.org) is developed by ChemAxon. Chemicalize uses ChemAxon’s name to structure parsing to identify chemical structures from web pages and other text sources. It provides a large variety of predicted data related to each structure [7].

**Figure 4.** Chemicalize—Link to ChemAxon's calculations of molecular properties.




## 4. What are the Differences between CWM Global Search and Other Systems such as PubChem or ChemSpider?

Unlike other systems such as PubChem or ChemSpider, CWM Global Search is NOT a database. We always search the most current snapshots of the supported data sources, thus also making sure that recently added records in the various data sources can be located.

This eliminates the problem that links to newly added PubChem records in ChemSpider and *vice versa* may not be found because of pending updates in the underlying database.

A major strength of CWM Global Search is the chemical structure search via the integrated structure editor (JDraw from Accelrys, Inc.). This structure editor supports copy/paste operations with major chemical drawing packages allowing the user to keep using his favorite structure drawing tool. Another unique feature of CWM Global Search allows you to draw a reaction; with one click you can find information for all reactants, reagents and products such as suppliers or safety information.

Comprehensive search—with a single click you can search for a structure, one or many associated CAS Registry Number(s), plus an arbitrary list of associated free text such as synonyms, brand names, and identifiers.

## 5. What are the Differences to SciFinder [8]?

SciFinder is the user interface for the world's largest chemistry database produced by the Chemical Abstract Service (CAS), a division of the American Chemical Society. However, it can only be accessed for a fee. While most academic institutions have access to SciFinder with an academic rate, many small and medium–sized companies simply cannot afford these fees, for them Global Search is a

very important first alternative. In all patent related cases, the most important question is whether a given structure is known or novel.

The most important aspect of the comparison with SciFinder is the fact that one cannot expect to find all answers to a given query in SciFinder. Many cases are known in whch the use of Global Search produced important references that were not found in SciFinder. Thus, Global Search will provide the occasional user with quick answers, and it will give the professional information specialist the very important extra certainty that his search was as comprehensive as possible. The strength of CWM Global Search is its access to additional sources that are not considered publications, like entries in Wikipedia and databases.

## 6. How to Start CWM Global Search?

You can start the application using the URL http://cwmglobalsearch.com/gsweb. If you start the application for the first time you might be asked to install the Microsoft Silverlight Plug-in. We support Internet Explorer, Firefox on Windows and Safari on Macintosh computers. Detailed information about the program can be found on the CWM Global Search homepage: www.akosgmbh.de/globalsearch. The free version is limited to Google, PubChem, ChemSpider, AKos Samples and the SureChem patent database. The free version will only show how many search results are found in the 40+ databases supported by the commercial version, and will not provide hyperlinks to the actual data.

## 7. Considerations

CWM Global Search relies on web services to generate the InChI names, and keys, and the availability of websites to search the underlying sources. We have no control over these web services and websites, they can be down for maintenance, moved to another location, or turned off. We periodically run searches to check for such issues, but a user should be aware that sometimes he has to re-execute a query when the web service or website is available again. According to our experience the availability fail rate of the web services and websites is very low. Since the whole application including the search engine is hosted on our server, we can upload a new version any time without involving the user, and we will do this because new interesting sources can be added monthly.

Maybe we should also discuss our business model. We try to keep the license fee as low as possible and augment this by giving sponsors room for their advertisements. A user in CWM Global Search will not interrupt his work to click on an advertisement, at least not very often. Therefore, the information that the sponsor wants to show must be in the advertisement. We run a slide show, giving each advertisement enough space to display an essential message. The picture will stay on the screen for a certain amount of time before the slide will change to a new one.

## References and Notes

1. From Wikipedia, the free encyclopedia. Available online: http://en.wikipedia.org/wiki/Metasearch_engine (accessed on 1 December 2010).
2. Heuser, U.J. *Denken, wie das Netz es will*; Die Zeit: Hamburg, Deutschalnd, 23 September 2010.

3. Carr, N. Is Google Making Us Stupid? Available online: http://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/6868/ (accessed on 1 December 2010).
4. Williams, A.J.; Tkachenko, V.; Lipinski, C.; Trophsa, A.; Ekins, S. Free online resources enabling crowd-sourced drug discovery. *Drug Discov. World* **2009**, *Winter*, 33–39.
5. Chemical Identifier Resolver beta 3. Available online: http://cactus.nci.nih.gov/chemical/structure (accessed on 1 December 2010).
6. JDraw is a Java applet structure editor from Accelrys, Inc.
7. What is chemicalize? Available online: http://www.chemicalize.org/about.php (accessed on 1 December 2010).
8. CAS Registry Number® and Synonym CAS Registry Number, SciFinder are registered trademarks of the American Chemical Society (ACS). All Rights Reserved.

## Appendix

**Table 1.** A list of the sources searchable in Global Search. An up-to-date list of data sources can be found at: http://www.akosgmbh.de/globalsearch/databases_in_gs.htm.

| Trademarks | Description | Link |
|---|---|---|
| *AKos Samples* | A database of approximate 6 million building blocks and screening compounds. All samples are checked for identity and purity by NMR. A network of suppliers can provide custom synthesis. | www.akosgmbh.de/AKosSamples |
| BASE Bielefeld Academic Search Engine | BASE is one of the world's most voluminous search engines, especially for academic open access web resources. BASE is operated by Bielefeld University Library. | www.base-search.net |
| BioMed Central The Open Access Publisher | BioMed Central is an STM (Science, Technology and Medicine) publisher which has pioneered the open access publishing model. | www.biomedcentral.com |
| BuyersGuideChem | BuyersGuideChem is a directory of chemicals and chemical suppliers on the Internet. | www.buyersguidechem.de |
| ChEBI | Chemical Entities of Biological Interest (ChEBI) is a freely available dictionary of molecular entities focusing on 'small' chemical compounds. | www.ebi.ac.uk/chebi |
| CHEMBANK | Initiative for Chemical Genetics. A freely available collection of data about small molecules (over 2000 compounds) and resources for studying their properties, especially their effects on biology. | http://chembank.broadinstitute.org |
| chemexper.com | The ChemExper Chemical Directory is mainly a supplier database for chemicals, and displays physical and chemical characteristics, structure, MSDS and more. | www.chemexper.com |
| Chemical BooK | A supplier database mainly for the Chinese market. | http://www.chemicalbook.com |

**Table 1.** *Cont.*

| | | |
|---|---|---|
| The Chemical Database | The Chemical Database will allow the user to retrieve information for any of 25,496 hazardous chemicals or 'generic' entries based on a keyword search. | http://ull.chemistry.uakron.edu/erd |
| CHEMICALLAND21 | Chemicalland21.com aims to be a resource of individual chemical information including technical data, safety data, and related compounds. | http://chemicalland21.com |
| ChemIDplus Advanced | This database allows users to search the NLM ChemIDplus database of over 370,000 chemicals. | http://chem.sis.nlm.nih.gov/chemidplus/ |
| | ChemSpider hosts the largest and most diverse online database of chemical structures sourced from over 150 different data sources. | http://www.chemspider.com/ |
| ChemSynthesis Chemical Database | ChemSynthesis is a database of compounds with their synthesis references and physical properties. | http://www.chemsynthesis.com/ |
| ClinicalTrials.gov | ClinicalTrials.gov is a registry of federally and privately supported clinical trials conducted in the United States and around the world. | http://clinicaltrials.gov/ |
| CiteXplore | ChEBI CiteXplore combines literature search with text mining tools for biology. | http://www.ebi.ac.uk/citexplore |
| ctd | Chemicals. CTD integrates a chemical subset of the Medical Subject Headings (MeSH®), the hierarchical vocabulary from the U.S. National Library of Medicine. | http://ctd.mdibl.org/ |
| DrugBank | 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. | http://www.drugbank.ca/ |
| DOAJ DIRECTORY OF OPEN ACCESS JOURNALS | The Directory of Open Access Journals (DOAJ) lists open access journals, that is, scientific and scholarly journals that meet high quality standards by exercising peer review or editorial quality control. | http://www.doaj.org |
| | Envirofacts contains chemical data from several different program system databases: the Aerometric Information Retrieval System, the Permit Compliance System, and the Toxics Release Inventory System. | http://www.epa.http://www.epa.gov/envirofw/gov/enviro fw/ |
| eMolecules | Find Suppliers and Information for over 8 million unique chemicals! | http://www.emolecules.com/ |

**Table 1.** *Cont.*

| | | |
|---|---|---|
| EPA | US Environmental Protection Agency | http://www.epa.gov/ |
| espacenet | This data source searches the European Patent Office database via the ChEBI CiteXplore search engine. | http://www.epo.org |
| FDA U.S. Food and Drug Admin Protecting and Promoting Your Healt | The FDA is responsible for protecting the public health by assuring the safety, efficacy, and security of human and veterinary drugs, biological products, medical devices, our nation's food supply, cosmetics, and products that emit radiation. | http://www.fda.gov |
| freepatentsonline all the inventions of mankind | Free patents online has hundreds of gigabytes of full-text data which is keyword searchable using the most powerful search engine in the industry. | http://www.freepatentsonline.com/ |
| Google scholar beta | Google Scholar is a freely-accessible Web search engine that indexes the full text of scholarly literature across an array of publishing formats and disciplines. | http://scholar.google.de/ |
| INCHEM | IPCS INCHEM is an invaluable tool for those concerned with chemical safety and the sound management of chemicals. | http://www.inchem.org/ |
| KEGG | KEGG COMPOUND is a chemical structure database for metabolic compounds and other chemical substances that are relevant to biological systems. | http://www.genome.jp/kegg/compound/ |
| NATIONAL CANCER INSTITUTE | The NCI database contains 250,251 structures, which corresponds to the open part of the NCI database up until and including the latest release of the DTP cancer screen results of August 2000. | http://129.43.27.140/ncidb2/ |
| Chemical Structure Lookup | Look up whether a structure occurs in many different databases, both public and commercial. Currently loaded pointers to over 74 million entries from more than 100 databases, representing more than 46 million unique chemical structures. | http://cactus.nci.nih.gov/cgi-bin/lookup/search |
| NEXTBIO | NextBio's integrated database contains publicly available data from a variety of sources, including GEO, caBIG, and Array Express among others. | www.nextbio.com |
| | National Institute of Allergy and Infectious Diseases: This database contains compounds that have been tested against HIV, HIV enzymes or opportunistic pathogens. | http://chemdb2.niaid.nih.gov |
| NIST Chemistry WebBook | The NIST Chemistry WebBook provides access to data compiled and distributed by NIST under the Standard Reference Data Program. | http://webbook.nist.gov/chemistry/ |

**Table 1.** *Cont.*

| | | |
|---|---|---|
| novo\|seek | novo\|seek is a biomedical search engine developed by bioalma for searching the published knowledge in biomedical literature. | http://www.novoseek.com |
| Open J-Gate | Open J-Gate is an electronic gateway to global journal literature in open access domain. | http://www.openj-gate.com |
| PharmGKB Pharmacogenomics Knowledge Base | Database of human genetic variations on drug response. | http://www.pharmgkb.org/ |
| PNAS | Proceedings of the National Academy of Sciences of the United States of America. PNAS Online contains the full text, figures, tables, equations, and references of all articles in PNAS dating back to 1990. | http://www.pnas.org/ |
| PubChem Structure Search | PubChem provides information on the biological activities of small molecules. | http://pubchem.ncbi.nlm.nih.gov/search/search.cgi |
| PubMed.gov | PubMed is a service of the U.S. National Library of Medicine that includes over 18 million citations from MEDLINE and other life science journals for biomedical articles back to 1948. PubMed includes links to full text articles and other related resources. | http://www.ncbi.nlm.nih.gov/pubmed/ |
| PubMed Central | PubMed Central (PMC) is a digital archive of life sciences journal literature that includes more than one million articles. | http://www.ncbi.nlm.nih.gov/pmc/ |
| *--SIRI MSDS Index—* | The SIRI MSDS archive is maintained by Dan Woodard, MD dan@siri.org and Ralph Stuart, CIH. Our objective is to make critical safety information immediately and universally as accessible as possible. | http://hazard.com/msds/ |
| SureChem | SureChem is making patent chemistry search faster, easier and more accessible. | http://www.surechem.org/ |
| Wikipedia The Free Encyclopedia | Wikipedia's articles provide links to guide the user to related pages with additional information. There are currently more than 5000 molecules indexed in Wikipedia. | http://www.wikipedia.org/ |
| ZINC | Database of commercially-available compounds for virtual screening. | http://zinc.docking.org/ |