

## How is an atom deleted from an AtomContainer?

You can remove a single atom with the method `AtomContainer.removeAtom(Atom)`. In contrast, the method `AtomContainer.removeAtomAndConnectedElectronContainers(Atom)` deletes an atom, its bonds and the lone pairs that are attached to it.

## How does the fingerprint algorithm work?

The hashed fingerprint algorithm implemented by CDK closely follows the approach taken by Daylight [2]. The algorithm generates all atom sequences up to 6 atoms that are present in a molecule by a depth-first search. Such a sequence might be a single atom (path length 1) or a longer chain of atoms like C-C-O-C=O. Then it generates a hash code for each sequence using the build-in hash function of the Java programming language. These hash codes initialize a pseudo-random number generator (RNG). The first random number between zero and 1023 generated by the RNG is taken to set the respective bit in the fingerprint. Due to the hashing function it is not guaranteed that different molecules map to different fingerprints. This is unavoidable with the myriads of possible molecular structures that are mapped on a fingerprint with only 1024 bits. Nevertheless, the algorithms' ability to distinguish between different structures is known to be quite good [3].

# Predictor

**Predictor is a stand-alone tool for predicting  $^{13}\text{C}$ -NMR shifts, based on NMRShiftDB data.**

by Stefan Kuhm

## Introduction

NMRSHIFTDB ([www.nmrshiftdb.org](http://www.nmrshiftdb.org), [1]) is a CDK-based project. It is an open-content web-database of nuclear magnetic resonance (NMR) data. One of its features is HOSE-code based spectrum prediction, which is available via the online interface. In order to be able to integrate this into offline applications, we now offer a stand-alone predictor, which can be downloaded at <http://www.nmrshiftdb.org/download/NmrshiftdbServlet/nmrshiftdb.sdf.zip?nmrshiftdbaction=predictor>.

## Usage

The downloaded JAR file contains a java class for performing the prediction and a dump of the current data. Additionally you need the follow-

## How is chirality handled in CDK?

Bond-based stereochemistry information can be read, e.g. from a MDL input file, and stored in the class **Bond**. **AtomParity** implements a storage container for atom-based stereochemistry. Unfortunately, at the time of this writing the atom-based stereochemistry information is not yet regarded by the SMILES parser.

Uli Fechner

Goethe-University Frankfurt, Germany

[u.fechner@chemie.uni-frankfurt.de](mailto:u.fechner@chemie.uni-frankfurt.de)

## Bibliography

- [1] David Weininger, Arthur Weininger, and Joseph L. Weininger. SMILES 2. Algorithm for Generation of Unique SMILES Notation. *J. Chem. Inf. Comput. Sci.*, 29:97–101, 1989.
- [2] Daylight theory manual. <http://www.daylight.com/dayhtml/doc/theory/theory.toc.html>, January 2004.
- [3] Robert D. Brown and Yvonne C. Martin. Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J. Chem. Inf. Comput. Sci.*, 36:572 – 584, 1996.

ing jars to run a prediction: `cdk-core.jar`, `cdk-extra.jar`, `JNL.jar`, all available from `nmrshiftdb` cvs at <http://cvs.sourceforge.net/viewcvs.py/nmrshiftdb/nmrshiftdb/lib/>. The prediction tool is easy to use: There is only one method, taking a molecule and one of its atoms as input and giving back the predicted values. Code could look like this:

```
MDLReader mdlreader = new MDLReader(  
    new FileReader(args[0])  
);  
Molecule mol = (Molecule)mdlreader.  
    read(new Molecule());  
PredictionTool predictor =  
    new PredictionTool();  
double[] result = predictor.  
    predict(mol, mol.getAtomAt(0));  
System.err.println(result[0] + " A "  
    + result[1] + " A " + result[2]);
```

The result is an array of doubles, the meanings are: 0=lower limit, 1=mean, 2=upper limit calculated via confidence limits, 3=median, 4=used spheres, 5=number of values, 6=standard deviation, 7=min value, 8=max value. Please note that the java vm should be started with at least 128 MB of mem-

ory, since data are read to memory in the constructor. This enables fast predictions once the object is created, but means you should make sure PredictionTool can be garbage collected when no longer needed.

An example of using Predictor would be to predict the  $^{13}\text{C}$ -spectrum of Aspirin (Fig. 1). With the data as of October 11, we get the following values:

Atom No.	Prediction	Spheres	SDBS
1	132.13	4	132.51
2	121.13	5	126.17
3	131.56	4	134.9
4	112.78	3	124.01
5	156.98	2	151.28
6	115.01	3	122.26
7	165.93	4	170.2
8	169.64	4	169.76
9	20.74	5	20.99

Reference values are from SDBS ([http://www.aist.go.jp/RIODE/SDBS/sdbs/owa/sdbs\\_sea.cre\\_frame\\_disp?sdbno=532](http://www.aist.go.jp/RIODE/SDBS/sdbs/owa/sdbs_sea.cre_frame_disp?sdbno=532)). Aspirin is not yet in

NMRShiftDB, as one can see from the sphere numbers; prediction would be the contained spectrum in case there was one.

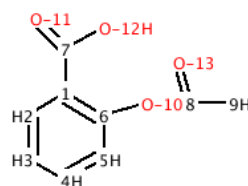


Figure 1: Structure of Aspirin.

Stefan Kuhn  
Universität zu Köln, Germany  
stefan.kuhn@uni-koeln.de

## Bibliography

- [1] C. Steinbeck. NMRShiftDB. *CDK News*, 1(1):2–3, 2004.

# Konqueror web shortcuts to the CDK API

This short article shows how a web shortcut can be defined in Konqueror that allows quick access to the CDK Application Programming Interface (API).

by Egon Willighagen

## Web shortcuts

Web shortcuts are a technology used in KDE (<http://www.kde.org>) that allow quick access to certain resources. For example, KDE 3.3 has the shortcut `dict:` defined that will look up a word in the Merriam-Webster Online ([www.m-w.com](http://www.m-w.com)). Some examples:

**dict:atom** Looks up *atom* in the Merriam-Webster Online

**gg:acetic acid site:woc.sci.kun.nl** Looks up *acetic acid* on <http://www.woc.sci.kun.nl/>

**db:orderedlist** Looks up the `<orderedlist>` element in the DocBook documentation

**doi:10.1021/ci025584y** Looks up the CDK article in the Journal of Chemical Information and Computer Sciences.

Such shortcuts can be ideal for web sites on which you often look up information. For example, I have created a web shortcut to ChemFinder ([http://](http://chemfinder.camsoft.com/)

[chemfinder.camsoft.com/](http://chemfinder.camsoft.com/)) using the CAS registry number to retrieve information about chemicals. To search for formaldehyde I type `chemfinder:50-00-0` (see Figure 1).

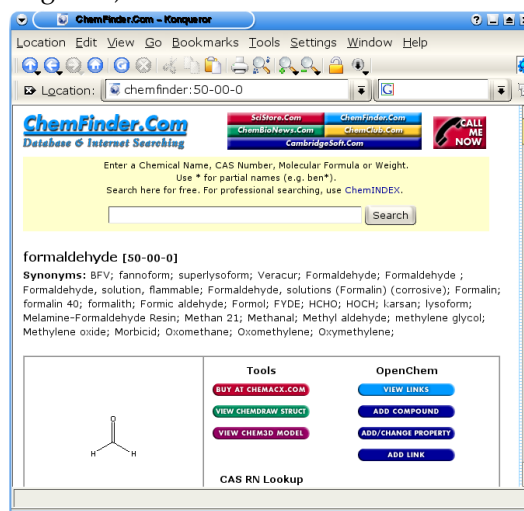


Figure 1: Web shortcut for ChemFinder.

## CDK API

The CDK API is created from the Java source code using the Javadoc utility. The API for the latest release is available on the CDK web site: <http://cdk.sf.net/api/>. You can then find the API for a spe-