The state of the CDK

Egon Willighagen
@egonwillighagen
0000-0001-7542-0286

2022-04-05, #cdk20y

Maastricht University

The history of the CDK

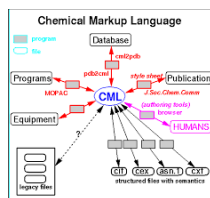**1997**

Chris Steinbeck

Peter Murray-Rust
Henry Rzepa

Maastricht University

Dear Christoph Steinbeck,

Yesterday I visited your site on JChemPaint. I like to contribute some of my expertise on Java and CML (1).

CML is a markup language that is able to contain chemical information. It can contain for example physical properties, for which I use CML in my Dictionary on Organic Chemistry (2). But is also might contain spectra, bibliographic references etc. And of course 2D and 3D structural information.

Therefore I propose to write both CML-input and -output procedures for the JChemPaint project.

I hope to hear from you soon.

Yours sincerely,

Egon Willighagen

1. http://www.xml-cml.org/
2. http://www.sci.kun.nl/sigma/Chemisch/Woordenboek/

https://chem-bla-ics.blogspot.com/2008/10/jchempaint-history-cml-patches-in-1999.html

Dear Egon,

thanks very much for your mail and your offer to write CML-input and
output routines for JChemPaint.
That really sounds great to me and I will give you access to our CVS
tree as soon as we have discussed the details.

Cheers,

Chris

--C. S.
Dr. Christoph Steinbeck (http://www.ice.mpg.de/~stein)
MPI of Chemical Ecology, Tatzendpromenade 1a, 07745 Jena, Germany
Tel: +49(0)3641 643644 - MoPho: +49(0)177 8236510 - Fax: +49(0)3641 643665

What is man but that lofty spirit - that sense of enterprise.
- Kirk, "I, Mudd," stardate 4513.3
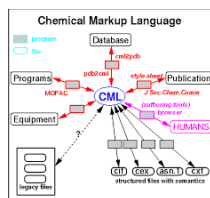
The history of the CDK

1997

Chris Steinbeck

Dan Gezelter

Peter Murray-Rust
Henry Rzepa

 Maastricht University

Maastricht University

The history of the CDK

1997

2000

Chris Steinbeck

Dan Gezelter

Peter Murray-Rust
Henry Rzepa

Maastricht University

The history of the CDK

2003

1997

2000

Chris Steinbeck

Dan Gezelter

Peter Murray-Rust
Henry Rzepa

Maastricht University

The history of the CDK

1997

JChemPaint

Chris Steinbeck

Jmol

Dan Gezelter

Chemical Markup Language

Peter Murray-Rust
Henry Rzepa

2000

CDK

2003

J. Chem. Inf. Comput. Sci. **2003,** *43,* 493–500

The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics

Christoph Steinbeck, Yongquan Han, Stefan Kuhn, Oliver Horlacher, Edgar Luttmann, and Egon Willighagen

Max-Planck-Institute of Chemical Ecology, Jena, Germany, TheraSTrat AG, Allschwil, Switzerland, Institute of Organic Chemistry, University of Paderborn, Germany, and Nijmegen, The Netherlands

Received August 17, 2002

2006

Current Pharmaceutical Design, 2006, *12,* 2111-2120

Recent Developments of the Chemistry Development Kit (CDK) - An Open-Source Java Library for Chemo- and Bioinformatics

Christoph Steinbeck, Christian Hoppe, Stefan Kuhn, Matteo Floris, Rajarshi Guha and Egon L. Willighagen

Cologne University Bioinformatics Center (CUBIC), Cologne, Germany; Pennsylvania State University, PA, USA and Institute for Molecules and Materials, Radboud University Nijmegen, The Netherlands

Maastricht University

The history of the CDK

1997

JChemPaint

Chris Steinbeck

Jmol

Dan Gezelter

Chemical Markup Language

Peter Murray-Rust
Henry Rzepa

2000

CDK

2003

2006

2017

Maastricht University

Three main CDK articles, and more...

Publications per year

Date	Work	Type	Topics
2022-03-03	Scaffold Generator - A Java library implementing molecular scaffold functionalities in the Chemistry Development Kit (CDK)	preprint	molecular scaffold // Chemistry Development Kit
2020-01-20	The rcdk and cluster R packages applied to drug candidate selection	scholarly article	R // drug discovery // Chemistry Development Kit
2017-06-06	The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching	scholarly article	cheminformatics // Chemistry Development Kit
2014-01-30	Efficient ring perception for the Chemistry Development Kit	scholarly article	ring // Chemistry Development Kit // chemical graph theory
2013-01-01	KNIME-CDK: Workflow-driven cheminformatics	scholarly article	cheminformatics // workflow // KNIME // Chemistry Development Kit
2013-01-01	Applications of the InChI in cheminformatics with the CDK and Bioclipse	scholarly article	Java // bioinformatics // cheminformatics // International Chemical Identifier // Bioclipse // Chemistry Development Kit
2012-02-02	LICSS - a chemical spreadsheet in microsoft excel	scholarly article	Microsoft // chemistry // Microsoft Excel // Chemistry Development Kit
2011-01-01	CDK-Taverna 2.0: migration and enhancements of an open-source pipelining solution	abstract	Taverna workbench // Chemistry Development Kit

An Open Science Cheminformatics World

Primary citations (from/to 3 CDK papers)

https://query.wikidata.org/#%23%20tool%3A%20scholia%0A%23defaultView%3ABarChart%0A%0ASELECT%0A%20%20%28STR%28%3Fyear_%29%20AS%20%3Fyear%29%0A%20%20%28SUM%28%3Fcount_%29%20AS%20%3Fcount%29%0A%20%20%3Fkind%0AWHERE%20%7B%0A%20%20%7B%0A%20%20%20%20VALUES%20%3Fyear_%20%7B%202000%202001%202002%202003%202004%202005%202006%202007%202008%202009%0A%20%20%20%20%20%20%20%20%20%20%20%20%202010%202011%202012%202013%202014%202015%202016%202017%202018%202019%20%7D%0A%20%20%20%20BIND%280%20AS%20%3Fcount_%29%0A%20%20%20%20BIND%28%22_%22%20AS%20%3Fkind%29%0A%20%20%7D%0A%20%20UNION%20%0A%20%20%7B%0A%20%20%20%20SELECT%0A%20%20%20%20%20%20%3Fyear_%0A%20%20%20%20%20%20%28COUNT%28DISTINCT%20%3Fciting_work%29%20AS%20%3Fcount_%29%0A%20%20%20%20%20%20%3Fkind%0A%20%20%20%20WHERE%20%7B%0A%20%20%20%20%20%20VALUES%20%3Ftarget%20%7B%20wd%3AQ27061829%20wd%3AQ27065423%20wd%3AQ30149558%20%7D%0A%20%20%20%20%20%20%3Fciting_work%20wdt%3AP2860%20%3Ftarget%20.%0A%20%20%0A%20%20%20%20%20%20%23%20Detect%20self-citations%0A%20%20%20%20%20%20BIND%28IF%28EXISTS%20%7B%20%3Ftarget%20wdt%3AP50%20%3Fselfauthor%20.%20%3Fciting_work%20wdt%3AP50%20%3Fselfauthor%20%7D%2C%0A%20%20%20%20%20%20%20%20%22detected%20incoming%20self-citations%22%2C%0A%20%20%20%20%20%20%20%20%22citations%20from%20others%20or%20non-detected%20self-citations%22%29%20AS%20%3Fkind%29%0A%0A%20%20%20%20%20%20%23%20Year%20of%20citation%0A%20%20%20%20%20%20%3Fciting_work%20wdt%3AP577%20%3Fdate%20.%0A%20%20%20%20%20%20BIND%28YEAR%28%3Fdate%29%20AS%20%3Fyear_%29%0A%20%20%20%20%20%20%7D%0A%20%20%20%20%20%20GROUP%20BY%20%3Fyear_%20%3Fkind%0A%20%20%20%20%7D%0A%20%20UNION%20%0A%20%20%7B%0A%20%20%20%20SELECT%0A%20%20%20%20%20%20%3Fyear_%0A%20%20%20%20%20%20%28COUNT%28DISTINCT%20%3Fcited_work%29%20AS%20%3Fcount_%29%0A%20%20%20%20%20%20VALUES%20%3Ftarget%20wdt%3AP2860%20%3Fcited_work%20.%0A%20%20%0A%20%20%20%20%20%20%23%20Detect%20self-citations%0A%20%20%20%20%20%20BIND%28IF%28EXISTS%20%7B%20%3Fcited_work%20wdt%3AP50%20%3Fselfauthor%20.%20%3Ftarget%20wdt%3AP50%20%3Fselfauthor%20%7D%2C%0A%20%20%20%20%20%20%20%20%22detected%20outgoing%20self-citations%22%2C%0A%20%20%20%20%20%20%20%20%22outgoing%20citations%20to%20others%20or%20non-detected%20self-citations%22%29%20AS%20%3Fkind%29%0A%0A%20%20%20%20%20%20%23%20Year%20of%20citation%0A%20%20%20%20%20%20%3Fcited_work%20wdt%3AP577%20%3Fdate%20.%0A%20%20%20%20%20%20BIND%28YEAR%28%3Fdate%29%20AS%20%3Fyear_%29%0A%20%20%20%20%20%20%7D%0A%20%20%20%20%20%20GROUP%20BY%20%3Fyear_%20%3Fkind%0A%20%20%20%20%7D%0A%7D%0AGROUP%20BY%20%3Fyear_%20%3Fkind%0AORDER%20BY%20DESC%28%3Fyear_%29

Citations to papers citing the CDK

citations	publication_date	citing_work	citing_workLabel
1293	2010-01-01	🔍 wd:Q27136473	MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data
880	2013-11-28	🔍 wd:Q27136674	STITCH 4: integration of protein-chemical interactions with user data
747	2014-01-01	🔍 wd:Q27136827	The IUPHAR/BPS Guide to PHARMACOLOGY: an expert-driven knowledgebase of drug targets and their ligands
742	2005-06-01	🔍 wd:Q27136404	Virtual Computational Chemistry Laboratory – Design and Description
717	2010-12-17	🔍 wd:Q27065420	PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints
590	2011-05-03	🔍 wd:Q27136894	SwissParam: A fast force field generation tool for small organic molecules
571	2007-01-01	🔍 wd:Q27136845	Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry
536	2006-02-22	🔍 wd:Q27062363	The Blue Obelisk-interoperability in chemical informatics
518	2016-01-29	🔍 wd:Q27702197	MetFrag relaunched: incorporating strategies beyond in silico fragmentation
496	2007-12-23	🔍 wd:Q27136964	SuperTarget and Matador: resources for exploring drug-target relationships
488	2018-03-19	🔍 wd:Q52647893	Extensive impact of non-antibiotic drugs on human gut bacteria.
418	2017-12-28	🔍 wd:Q48024105	Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks.
392	2004-03-01	🔍 wd:Q27136493	VEGA – An open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming
359	2004-01-01	🔍 wd:Q24599948	BRENDA, the enzyme database: updates and major new developments
336	2010-01-01	🔍 wd:Q21284354	In silico fragmentation for computer assisted identification of metabolite mass spectra
307	2007-12-23	🔍 wd:Q28465606	STITCH: interaction networks of chemicals and proteins
302	2011-01-01	🔍 wd:Q27137071	PREDICT: a method for inferring novel drug indications with application to personalized medicine

Maastricht University

https://w.wiki/4$y9

Tools using the CDK

	Recent release	CDK version	Comments
Java			
JChemPaint			Updated needed!
AMBIT		2.1.1	
Bioclipse	2006	1.3.5	Now Bacting
Bacting	March 2022	2.7.1	
Workflow systems			
KNIME	2021	1.5.3	
Taverna			
Squonk		2.7.1	
Python			
cinfony			
pybacting	March 2022	2.7.1	Uses Bacting
R			
rcdk	2021-10-17	2.3 (?)	

https://github.com/cdk/

https://github.com/cdk/

- Git (after CVS and SVN)
- Maven (after Ant)

- GitHub and SourceForge
- Peer review (since 2005?)
- DOI (via GitHub/Zenodo)

- 118 authors

GitHub / Zenodo

CDK 2.7.1 (Latest)   Compare ▾   ✎ 🗑

👤 **johnmay** released this Jan 11, 2022   · **227 commits** to master since this release   🏷 cdk-2.7.1   ⊶ 8d5821f

---

`DOI` `10.5281/zenodo.5837566`

This page documents the changes for CDK v2.7 and v2.7.1. The patch version was made after some minor issues with how the new InChI code was organised were discovered by downstream projects.

Features

Switch from JNI to JNA InChI.

There are two main technologies for calling native code JNI (Java Native Interface) and JNA (Java Native Access). JNI requires writing a custom native wrapper which is then bound to Java code, JNA allows you to call the native methods of an existing SO/DYLIB directly. Essentially what this means is to expose the native InChI library in Java one needs to first write (and maintain) a native wrapper, with JNA we can just drop the InChI SO directly in. JNI InChI exposed InChI v1.03 and worked well for many years - unfortunately this project was no longer maintained and as newer more stable versions of InChI were released (now v1.06) an alternative was needed. A few years ago Daniel Lowe started JNA InChI and recently made it feature complete and released v1.0.

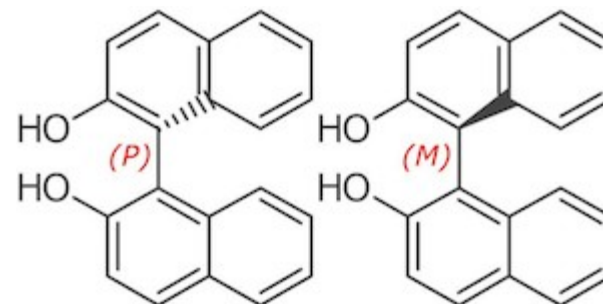ChemAxon have also independently used the JNA path to integrated newer InChI libraries into their tools: (slides). It is not clear if this was made available, it is not listed on GitHub/ChemAxon.

Build on Java 17

The Maven plugins were updated to allow building on Java 17

Maastricht University

CDK 2.1 (2017-12)

- IAtom.getbonds()
- Stereochemistry API

```
114 John Mayfield
 25 Kazuya Ujihara
  5 Horlacher, Oliver
  4 Egon Willighagen
  2 Noel O'Boyle
  2 Jakub Galgonek
  2 Hirotomo Moriwaki
  1 Tomáš Pluskal
  1 Sebastian Fritsch
  1 Christoph Steinbeck
```

Maastricht University
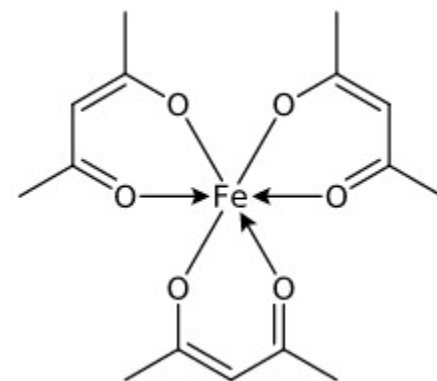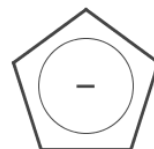
CDK 2.2 (2018-10)

- SMARTS engine
- IsotopePatternGenerator
- DfPattern for substructure search
- Many other incremental changes

```
213 John Mayfield
 35 Saulius Gražulis
 17 Kazuya Ujihara
  9 bach
  8 Egon Willighagen
  6 michaelwenk
  4 Rajarshi Guha
  4 Lyle D. Burgoon
  2 Katrin Leinweber
  1 potatostodie
  1 meier-rene
  1 balaji
  1 Sundaram
  1 Raven
  1 Matt Swain
  1 Martin Gütlein
```
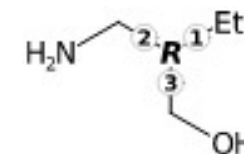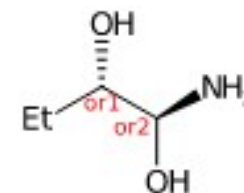
CDK 2.3 (2019-09)

- JPLogP (Lhasa)
- fixed ALogP
- donuts are back
- other rendering options

```
163 John Mayfield
 22 Egon Willighagen
 10 Kazuya Ujihara
  6 ficolas2
  5 Jeffrey Plante
  2 Markus Sitzmann
  1 martin morissette
  1 Qinqing Liu
```

CDK 2.5 (2021-05)

- IElement by atomic number
- Racemic/Relative stereo groups
- MDL Data Sgroups
- more rendering goodies


```
169   John Mayfield
 75   Egon Willighagen
 24   Danny Katzel
 18   Xavier Linn
 11   Mark J. Williamson
  9   Mark Williamson
  2   Jeffrey Plante
  2   Kazuya Ujihara
  1   Tagir Valeev
  1   Jean Marois
  1   Kai Dührkop
  1   Glur, Marco
  1   Robin Schmid
```

CDK 2.6 (2021-12)

- Log4j 2.15
- Bunch of fixes

```
18 John Mayfield
 9 Egon Willighagen
 1 Danny Katzel
 1 AndusDEV
```

Maastricht University

CDK 2.7.1 (2022-01)

- more log4j
- from JNI to JNA InChI
- Java 17
- Remove Guava
- CMLXOM

```
137 John Mayfield
  6 Egon Willighagen
  1 dependabot[bot]
```

The state of the CDK

	Recent release	CDK version	Comments
Java			
JChemPaint			Updated needed!
AMBIT		2.1.1	
Bioclipse	2006	1.3.5	Now Bacting
Bacting	March 2022	2.7.1	
Workflow systems			
KNIME	2021	1.5.3	
Taverna			
Squonk		2.7.1	
Python			
cinfony			
pybacting	March 2022	2.7.1	Uses Bacting
R			
rcdk	2021-10-17	2.3 (?)	

The state of the CDK

- 1. command line interface

- 2. Python API

- 3. more unit testing (test coverage)

- 4. updated tools

- 5. bench marks against X

- 6. CDK News (or just more blogging about it)

- 7. <put your wishlist here>

Maastricht University

The state of the CDK

Egon Willighagen
@egonwillighagen
0000-0001-7542-0286

2022-04-05, #cdk20y

Maastricht University