

The state of the CDK

Egon Willighagen
@egonwillighagen
0000-0001-7542-0286

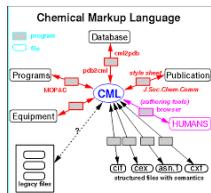
2022-04-05, #cdk20y
CC-BY, doi:10.5281/zenodo.6414204

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 it 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. Maastricht University <https://chem-bla-ics.blogspot.com/2008/10/jchempaint-history-cml-patches-in-1999.html>
<http://www.sci.kun.nl/sigma/ChemischWoordenboek/>

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
Maastricht University

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

The history of the CDK

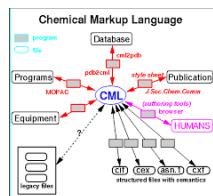
1997



Chris Steinbeck



Dan Gezelter



Peter Murray-Rust
Henry Rzepa

CHEMWorld (ChemSeq[], ChemModel[], measures[])

ElementFactory
AtomContainer
IsotopeContainer
AngleTorsion

Distance
Angle
Dihedral

BasisSet
IsoSurface
BondsFromResidues
WedgeFromResidues

BondFinder
BondFinder
TorsionFinder
RingFinder
RingFinder
SORFinder

MoleculeFinder
AAResidueFinder

DNASeqFinder

SemanticLink (ChemModel (set of the where []))

Solution

AtomType

```

ChemObject <-- ChemObjectListBase (fineChem)
Element (Symbol, at %, At. mass)
AtomType (Hybridization, radius, size, default # of bonds, symbol, max bond)
Isotope (mass, spin, symbol)
Atom (Element, point2D, point3D, hydrogen count, degree, parity, impreciseCount)
ElectronContainer (ElectronCount)
Bond (atoms[], degree, stereo, electron count, charge)
Orbital (Orbital, point3DSizes, point3D, magnitude)
Bend (atoms[])
Torsion (atoms[], :=Improper)
Path (atoms[])
Ring (atoms[], isAromatic)
Set-of-Rings (rings[])
Residue (atoms[], name)
DNAbase (atoms[], name)
Sequence (Residues[], phys[], pos[])
Molecule (atoms[], bonds[], bends[], torsions[], rings[], residues[], DNA)
Set of Molecules (molecules[])

```

ChemSequence (ChemModel[])
ChemFile

RENDERER

Atom selections

3DRadar (CPK/ball-stick/etc generation)

3DRaster Camera position

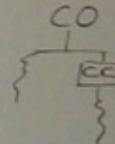
Atoms

Blank

Ar —> X —> Y —> Z —> C —> N —> O —> S —> F

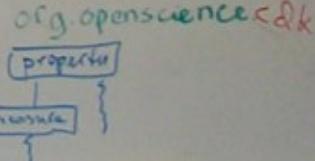
getAtoms
get Bonds
get Lines
get molecules

getAtoms
get Bonds
get Lines
get molecules



Input Bonds
Input Lines

Crystal
Quantum
Bloinfa
Interactions
measures
spectra



<http://www.steinbeck-molecular.de/steinblog/index.php/2009/06/26/historical-cdk-document-unearthed/>



The history of the CDK

1997

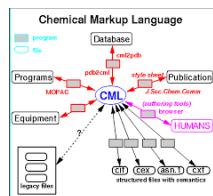
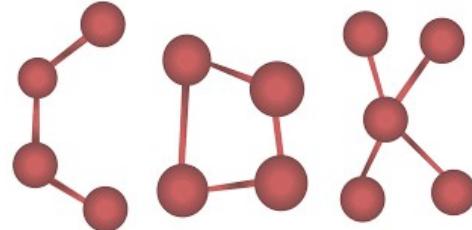


Chris Steinbeck

Jmol

Dan Gezelter

2000



Peter Murray-Rust
Henry Rzepa

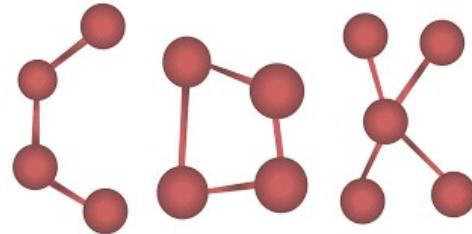
2003

The history of the CDK

1997



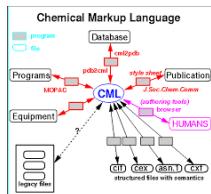
2000



Chris Steinbeck

Jmol

Dan Gezelter



Peter Murray-Rust
Henry Rzepa

Maastricht University

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

493

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

Christopher Steinbeck,^{a,*} Yongqian Han,^a Sufia Khan,^a Oliver Horlein,^b Edgar Lippmann,^c and Egwin Willighagen^d

^a Max Planck Institute of Chemical Energy Conversion, Düsseldorf, Germany; ^b Michael Schmidbauer, Institute of Organic Chemistry, University of Paderborn, Germany; and ^c Nijmegen, The Netherlands

Received August 17, 2002

The Chemistry Development Kit (CDK) is a freely available open-source Java library for Structural Chemistry and Bioinformatics. Its architecture and capabilities as well as development as an open-source project. The CDK is designed to support the needs of chemists and bioinformaticians in their work. The CDK provides methods for many common tasks in molecular informatics, including 2D and 3D rendering of molecules, chemical reaction planning, molecular properties, pharmacophore generation, fragment generation, and potential energy minimization. Application scenarios as well as source information for interested users and potential contributors are given.

1. INTRODUCTION

Whether pursue the endeavor of creating a larger software package in bioinformatics or continue the tradition of writing small programs to solve specific problems, the lack of implementing the standard representation of chemical structures in a consistent manner remains the problem for many chemists. In this paper, the problem of commonly available cheminformatics libraries is discussed. The authors are from the MEL Informatics System, Inc., Daylight Chemical Information Systems, CambridgeSoft, Cambridge, MA, and certainly many others. A scientist is an academic, and academic research is the primary motivation for his results with the scientific community. Using propriety software, however, is not always the best way to do science. In fact, scientific software is too often chosen, leaving the user with a black box performing magical operations without understanding what is really going on overall scientific progress. By publishing how one can implement a standard representation of chemical structures, thereby advancing the practical field as a whole, we hope to encourage the use of open-source software and, hopefully, reduce the cost of software packages.

Most importantly, if the community of users is large enough, the software will be used by many people, and they should not take too long until a practical software is developed. As Raymond E. Leeputten said, "If there are shelves," a few Raymond put it in his widely recognized book, "The Art of Computer Programming," "there is no need to do so much." In other words, the scientific community should be given the opportunity to contribute to the development of the mechanisms and the principles of the open-source movement. Good can still give rise to appropriate form, because the source code is given to the community, the package is a whole system, and the user can change the source code to fit his needs and the nature of the license. This copyright notice must not be removed from the source code. If changes and improvements made by others can, of course, be published, the original author of the code should be acknowledged. We believe that the copyright holder will easily accept this, as he will be able to benefit from the improvements made by others. This is especially important for academic research, as the scientific community is the primary user of open-source software in a particular field.

We have developed a series of open-source software on one hand and the scientific tradition on the other hand, we stated "We are SourceForge," "With hard-open-source software, we can benefit from the contributions of many people from about 100 developers from about five different countries. It is a great example of how different people can work together to produce something that is useful to the scientific community."

*Corresponding author present address: CambridgeSoft, Cambridge, MA 02142, USA. E-mail: csteinbeck@mpic.de

† To whom reprint requests should be addressed.

© 2003 American Chemical Society

10.1021/ci020259q CCC: \$25.00 © 2003 American Chemical Society

Published on Web 02/13/2003

2003

The history of the CDK

1997



2000

J. Chem. Inf. Comput. Sci. 2003, 43, 493–500
The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemoinformatics and Bioinformatics

493

Cheng Steinbeck,^{a,*} Yongyan Han,^a Stefan Kuhn,^a Oliver Horlein,^b Edgar Lippisch,^c and Egon Willighagen^d

^a Max-Planck-Institute of Chemical Energy Conversion, Duvelstraat 4, Münster, Germany; ^b Institute of Organic Chemistry, University of Paderborn, and Nijmegen, The Netherlands

Received August 17, 2002

The Chemistry Development Kit (CDK) is a freely available open-source Java library for Structural Chemoinformatics and Bioinformatics. Its architecture and capabilities as well as its development as an open-source project are described. The CDK provides methods for many common tasks in molecular informatics, including 2D and 3D rendering of molecules, generation of chemical structures, SMILES parsing and conversion, reaction searching and planning, fragment generation, Application scenarios as well as access information for interested users and potential contributors are given.

1. INTRODUCTION

Wherever persons the endeavor of creating a larger software package in chemoinformatics or bioinformatics, they will face the same problem as we did: the complete lack of implementing the standard representation of chemical structures in a way that is compatible with the needs of the scientific community. This is because there is no common standard representation of chemical structures that can be used by all.

Most importantly, if the community of users is large enough, it is likely that the software will be used by others than the original developer. In this case, the developer should not take too long until a practical software is released. "It is not good to have a software that sits on shelves," says Eric Raymond put it in his widely recognized book "The Cathedral and the Bazaar".¹ He means that the mechanisms and principles of the open source movement should be applied to the development of the software.

Good can and should be given for appropriate fees, because the developer is entitled to compensation for his work. In other words, scientific software is often offered source licensing the user with a "Black box" performing magical operations. This is not good for the scientific community or overall scientific progress, this trend formally seems to be accepted by the scientific community. It is also probably the most central paradigm in science. By publishing their results, scientists make them available to the public upon their results, thereby advancing the partial field as a whole. This is the reason why the scientific community is full of grants, "Issue Novelties". One of the motivations for such commercial software is to earn money. This is understandable, the price of scientific recognition and reputation among peers is high.

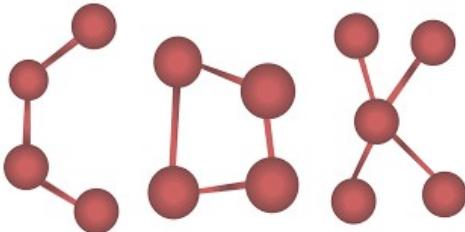
In the recent years the idea sketched above has been part of the open source movement. The term "open source" is not well known, but the concept is well known and the nature of the license. This copyright notice must not be removed from the code. The copyright notice must be included in every file and every class. All changes and improvements made by others can, of course, be published under the same license. This is important for the developer, because he can give his code to anyone else for free. This is also true for the copyright holder will easily accept this. The developer can also publish his code under a different copyright license. This is especially important for academic institutions, because they can publish their own developments in a particular field.

There are two main areas of open-source software: on one hand and the scientific tradition on the other hand, we stated above. The scientific tradition is represented by the "Apache License". We are "SunForge", a "With broad open-source licenses, SunForge is the first company to offer open-source products from about 10 developers from about five different countries. Apache License is a trademark of the Apache Software Foundation. © 2003 American Chemical Society
Published on Web 02/13/2003

Chris Steinbeck

Jmol

Dan Gezelter



2006

Curr Opin Pharmacol, 2006, 12, 2101–2120

2111

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

Christoph Steinbeck,^{a,*} Christian Hoppe,^a Stefan Kuhn,^a Matteo Floris,^a Rajarsi Guha^b and Egon L. Willighagen^c

^a Cologne University Biocomputing Center (CUBIC), Cologne, Germany; ^b Pennsylvania State University, PA, USA and ^c Institute of Chemical Biology and Drug Design, University of Cologne, Cologne, Germany

Abstract: The Chemistry Development Kit (CDK) provides methods for common tasks in molecular informatics, including 2D and 3D rendering of chemical structures, 3D rotamers, SMILES parsing and conversion, reaction searching, reaction planning, fragment generation, and the recently introduced interface to statistical software.

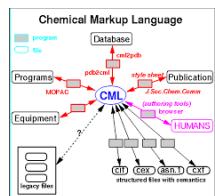
1. INTRODUCTION

From the very beginning, there was an understanding that chemoinformatics is a scientific discipline, which requires the use of computers in chemistry with methods developed in computer science. This rather broad definition by Julianus Böhlke and Hans-Joachim Wilen in 1993 originates from a diverse set of fields – including mathematics, statistics, computer science, and chemistry. Machine learning, neural networks, and other statistical methods are applied to create, process and understand chemical structures. In addition, chemoinformatics methods include characterization of chemical structures using graph theory, and the use of neural networks and other statistical methods and approaches to predict properties of chemical structures. With the advent of high throughput methods such as high throughput screening and combinatorial chemistry, the high demand of processing speed as well as greater accuracy in prediction has led to the development of new chemoinformatics methods and tools have increased.

Though in the last 20 years, many research groups have been devoted to the study of chemoinformatics, only a few have been able to develop a truly useful software package to support even the most common tasks in chemoinformatics. This is mainly due to the fact that the field of chemoinformatics is very interdisciplinary. The enormous progress made in the molecular sciences over the last 20 years has led to the fact that the chemoinformatics projects have only been possible because it was supported by a biocomputing software culture of openness.

In the academic community a lot of research has been done in isolated areas of chemoinformatics, but for a long time no general software package was available. The first publicly available software package to support even the most common tasks in chemoinformatics was the Chem3D Pro, developed by CambridgeSoft. The enormous progress made in the molecular sciences over the last 20 years has led to the fact that the chemoinformatics projects have only been possible because it was supported by a biocomputing software culture of openness.

The function of a chemoinformatics toolkit, by definition, is to represent, generate and process chemical structures. These are sets of numeric values that mathematical models can use to predict properties of chemical molecules. Molecular descriptors are used in a number of areas within chemoinformatics. In the early days of chemoinformatics, one line of work in the CDK project has been to focus on the use of molecular descriptors for reaction planning and QM3 modeling environments. To this end a number of molecular descriptors have been implemented in the CDK. Table I gives an overview of descriptors currently implemented in the CDK. The section describes the general design of the descriptor package.



Peter Murray-Rust
Henry Rzepa



Maastricht University

*Address correspondence to the author at the Cologne University Biocomputing Center (CUBIC), University of Cologne, Cologne, Germany. Tel.: (0221) 470-7786. E-mail: c.steinbeck@kub.uni-koeln.de

© 2006 Blackwell Science Publishers Ltd.

1361-4128/06 \$30.00.00

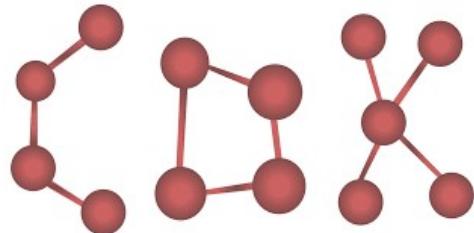
2003

The history of the CDK

1997



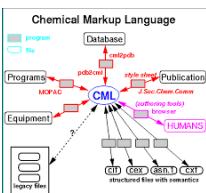
2000



Chris Steinbeck

Jmol

Dan Gezelter



Peter Murray-Rust
Henry Rzepa

Maastricht University

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

493

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

Christoph Steinbeck,^{a,*} Yongyan Han,^a Stefan Kuhn,^b Oliver Hechler,^c Edgar Lattman,^c and

Egon Willighagen^d

^a Max-Planck-Institute of Colloid and界面物理化学生物学研究所, 德国达姆施塔特; ^b Michael Schmidts, Institute of Organic Chemistry, University of Paderborn, Germany; and ^c Nijmegen, The Netherlands

Received August 17, 2002

The Chemistry Development Kit (CDK) is a freely available open-source Java library for Structural Chemoinformatics and Bioinformatics. Its architecture and capabilities as well as its development as an open-source project are described. The CDK provides methods for the manipulation of chemical structures, for the generation of 2D and 3D visualizations, for the calculation of physicochemical properties, for the generation of chemical reaction schemes, and for the generation of pharmacophores. Application scenarios as well as access information for interested users are given.

1. INTRODUCTION

Wherever persons the endeavor of creating a larger software package in chemoinformatics or bioinformatics, they will face the problem of how to deal with the typical lack of implementing the standard representation of chemical structures. In this paper, we describe the lack of commonality in molecular informatics that the CDK can currently accommodate (Chemoinformatics) and the lack of commonality in molecular bioinformatics (Bioinformatics). According to recent reports, the number of scientific publications that have been published in the field of chemoinformatics has increased dramatically over the last few years. MEL Informatics Systems, Inc., Daylight Chemical Information Systems, CambridgeSoft, CambridgeSoft, CambridgeSoft, and certainly many others. A scientist at an academic institution or a pharmaceutical company can hardly ignore his results with the scientific community. Using proprietary software, however, it is often difficult to publish his work to do so.

Most recently, scientific software has been chosen more often to work with a black box performing magical operations on the user's behalf. This is especially true for overall scientific projects, this trend formerly seems to be increasing. However, the use of black boxes is probably the most central paradigm in science. By publishing his results with the scientific community, the scientist can update his results, thereby advancing the partial field as a whole. This is the reason why the use of black boxes is not a good idea. Issue No. 1! One of the motivations for such communication is the desire for recognition and, of course, the desire for social recognition and reputation among his peers.

In the first part of this article, we sketch above how part of the present situation has come about and what can be done to improve it.

* Corresponding author. Current address: Computer Science Department, The University of Texas at Austin, TX 78712-0265, U.S.A.; e-mail: csteinbeck@mail.utexas.edu

† Present address: Department of Chemistry, University of Paderborn, Germany.

© 2003 American Chemical Society

Published on Web 02/13/2003

Abstract: The Chemistry Development Kit (CDK) provides methods for common tasks in molecular informatics, including 2D and 3D rendering of chemical structures, IO routines, SMILES parsing and printing, reg reactions, isomorphism checking, reaction diagram generation, etc. Implemented in Java, it is used both for server-based computational services, possibly running on a mainframe, and for desktop applications. The CDK is freely available under the terms of the GNU General Public License (GPL). We introduce the CDK and discuss its architecture, the new QM3L capabilities and the recently introduced interface to open-source software.

Keywords: Java, Chemoinformatics, Bioinformatics, Metabolomics, Description

2006

Curr Opin Pharmacol 2006, 12: 2101–2120

2111

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

Christoph Steinbeck^{a,*}, Christian Hoppe,^b Stefan Kuhn,^b Matteo Floris,^b Rajarsi Guha^b and Egon Willighagen^c

^a Cologne University Biocomputing Center (CUBIC), Cologne, Germany; ^b Pennsylvania State University, PA, USA and ^c Institute of Chemical Biology and Computing, Chinese Academy of Sciences, Shanghai, China

Abstract: The Chemistry Development Kit (CDK) provides methods for common tasks in molecular informatics, including 2D and 3D rendering of chemical structures, IO routines, SMILES parsing and printing, reg reactions, isomorphism checking, reaction diagram generation, etc. Implemented in Java, it is used both for server-based computational services, possibly running on a mainframe, and for desktop applications. The CDK is freely available under the terms of the GNU General Public License (GPL). We introduce the CDK and discuss its architecture, the new QM3L capabilities and the recently introduced interface to open-source software.

Keywords: Java, Chemoinformatics, Bioinformatics, Metabolomics, Description

2017

Wiley Interdiscipl Rev Cheminform 2017; 8:e1294

DOI: 10.1002/wicr.1294

Journal of Cheminformatics

Open Access

The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular descriptors, and substructure searching

Egon L. Willighagen^a, John W. Mayfield^a, Jonathan Alvisian^b, Arvid Berg^c, Lars Carlsson^c, Stefan Kuhn^b, Tomáš Pásek^b, Miguel Rojas-Chert^b, Olof Sjöström^b, Gillete Tononice^b, Chris T. Ebel^b, Rajarsi Guha^b and Christoph Steinbeck^b

Abstract

Background: The Chemistry Development Kit (CDK) is a widely used open-source chemoinformatics toolkit providing data structures to represent chemical concepts along with methods to manipulate them. The library implements a wide variety of chemoinformatics algorithms ranging from chemical structure representation to reaction search and planning, retrosynthesis analysis, pharmacophore modeling, and pharmacokinetic calculations. The CDK has grown significantly, however, resulting in many complex dependencies among its components and rendering its maintenance difficult.

Conclusions: This paper highlights our continued efforts to provide a community driven, open source chemoinformatics toolkit that is highly reliable, performant, and extensible. We introduce the CDK v2.0 since the last release, which includes the increased functionality and performance of the CDK since its initial release. We highlight the increased functionality and performance for substructure searching, molecular fingerprints, and rendering of molecules. Second, we outline how the CDK has evolved with respect to quality control and the approach we have adopted to ensure stability, including a code review system.

Keywords: Java, Chemoinformatics, Bioinformatics, Metabolomics, Description

Background: The CDK is a widely used open-source chemoinformatics toolkit providing data structures to represent chemical concepts along with methods to manipulate them. The library implements a wide variety of chemoinformatics algorithms ranging from chemical structure representation to reaction search and planning, retrosynthesis analysis, pharmacophore modeling, and pharmacokinetic calculations. The CDK has grown significantly, however, resulting in many complex dependencies among its components and rendering its maintenance difficult.

Conclusions: This paper highlights our continued efforts to provide a community driven, open source chemoinformatics toolkit that is highly reliable, performant, and extensible. We introduce the CDK v2.0 since the last release, which includes the increased functionality and performance of the CDK since its initial release. We highlight the increased functionality and performance for substructure searching, molecular fingerprints, and rendering of molecules. Second, we outline how the CDK has evolved with respect to quality control and the approach we have adopted to ensure stability, including a code review system.

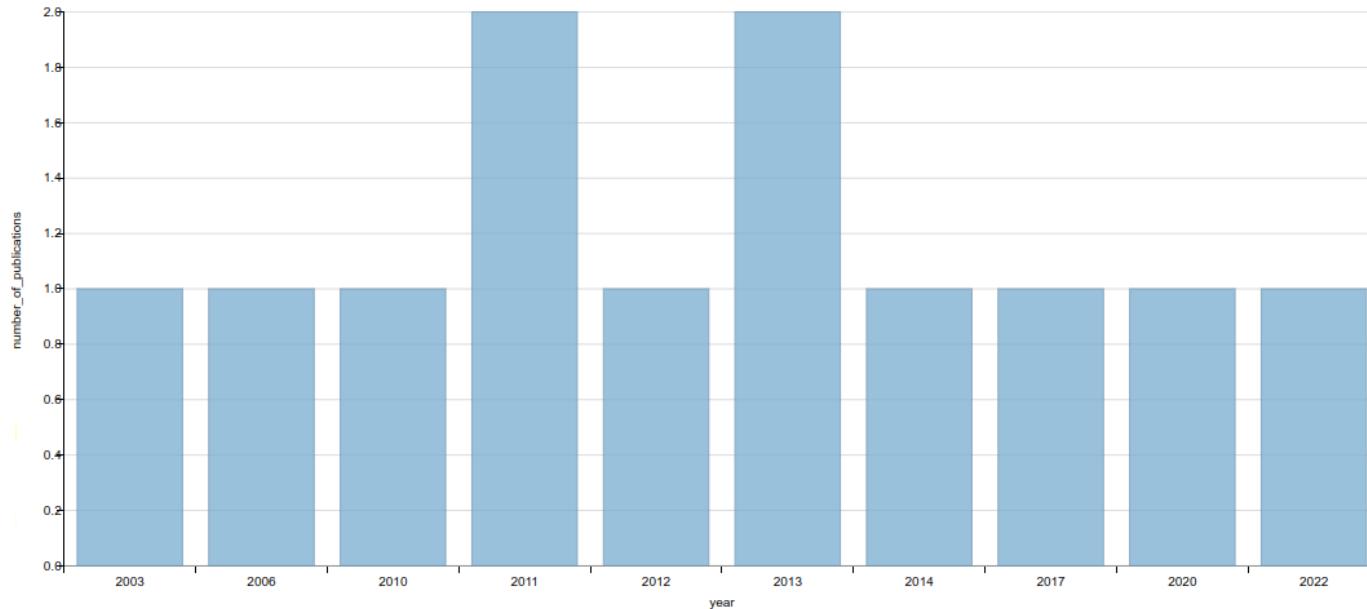
Keywords: Java, Chemoinformatics, Bioinformatics, Metabolomics, Description

^a Department of Agro-Environmental Engineering, Ghent University, Belgium; ^b The Netherlands Institute for Space Research (NIR), Delft, The Netherlands; ^c KNIME, Berlin, Germany

^a Springer Open

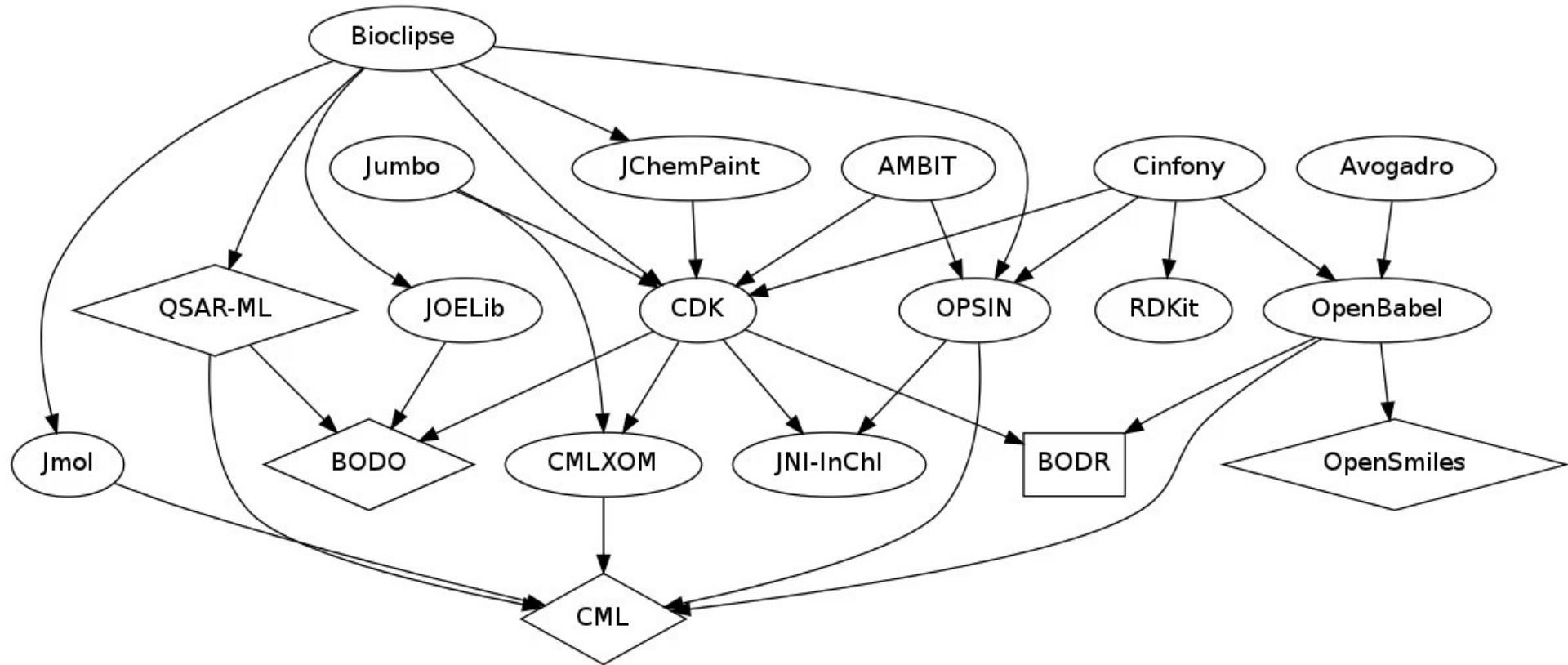
Three main CDK articles, and more...

Publications per year

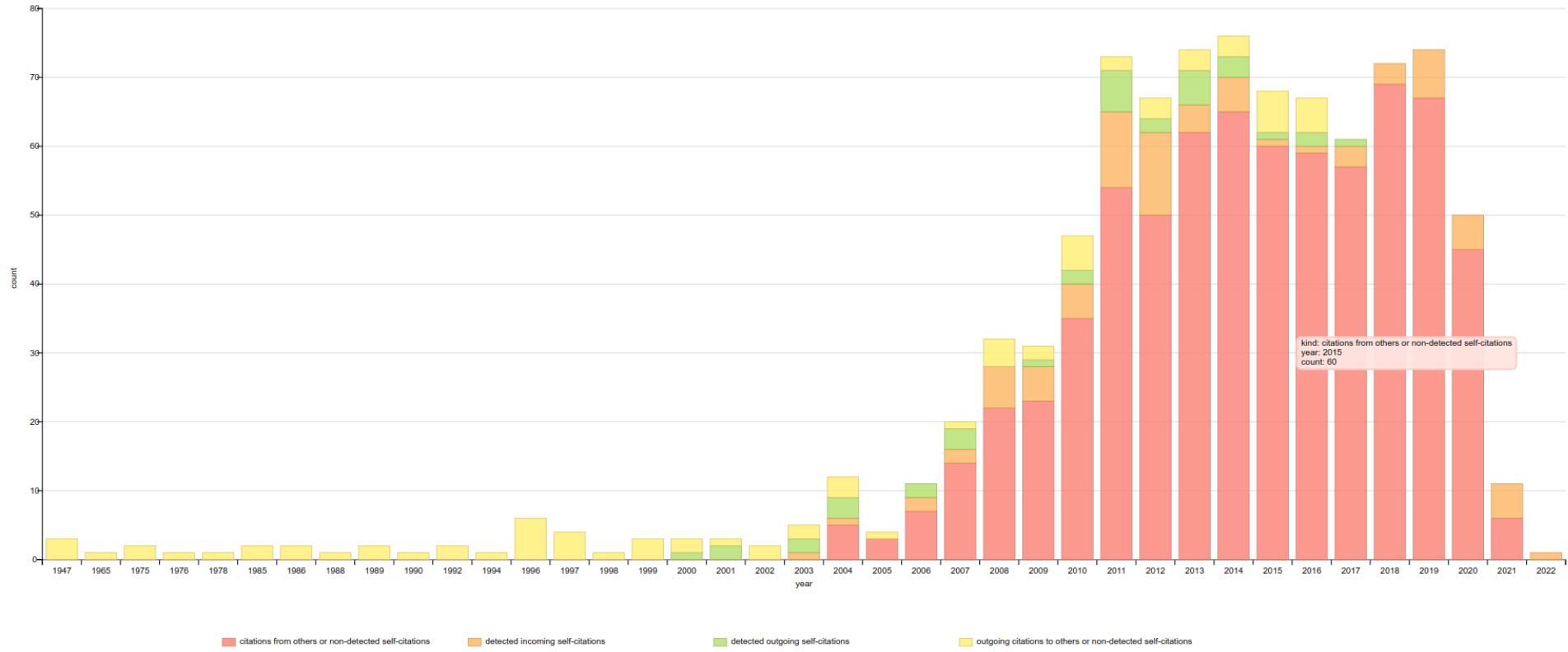


Date	Work	Type	Topics
2022-03-03	Scaffold Generator - A Java library implementing molecular scaffold functionalities in the Chemistry Development Kit (CDK)	preprint	molecular scaffold // Chemistry Development Kit
2020-01-20	The rcdk and cluster R packages applied to drug candidate selection	scholarly 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)



Citations to papers citing the CDK

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

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/>

master ▾ 18 branches 73 tags Go to file Add file ▾ Code ▾

johnmay Merge pull request #849 from egonw/update/anotherone ...	88447fc 10 days ago	17,152 commits
.github/workflows	Enable Jacoco for SonarCloud	last month
app	First batch of NPE checks	last month
base	Fix JavaDoc errors due to copy-paste	last month
bundle	Dev version 2.8-SNAPSHOT ready for changes.	3 months ago
descriptor	javadoc improvements for SmallRingDescriptor	16 days ago
display	Need float here.	last month
doc	Depth-first (DF) substructure search algorithm.	4 years ago
legacy	Misc NoSuchElementException needed on these iterators.	last month
misc	Don't ignore important return values - the iterators were actually ok...	last month
src/site/apt	Site index page	8 years ago
storage	allows both space and comma as separator for InChI options	28 days ago
tool	We only need volatile if something is likely and inefficient to be ac...	last month
.gitignore	Added PaDEL AtomPairs2D fingerprint.	4 years ago
AUTHORS.txt	Two new contributors	16 days ago
ISSUES.txt	More standardisation.	4 years ago
LICENSE.txt	More standardisation.	4 years ago
README.md	Update README.md	last month
checkstyle.xml	Merge pull request #185 from egonw/fix/missingJavadoc	6 years ago
codecov.yml	Don't post comments on PRs	15 months ago

About The Chemistry Development Kit
cdk.github.io/
java bioinformatics chemistry
cheminformatics code4lib blueobelisk

Readme
LGPL-2.1 License
336 stars
37 watching
144 forks

Releases 18
CDK 2.7.1 Latest on Jan 11 + 17 releases

Used by 15

Contributors 65 + 7

+ 54 contributors

<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

-o 8d5821f

DOI [10.5281/zenodo.5837566](https://doi.org/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

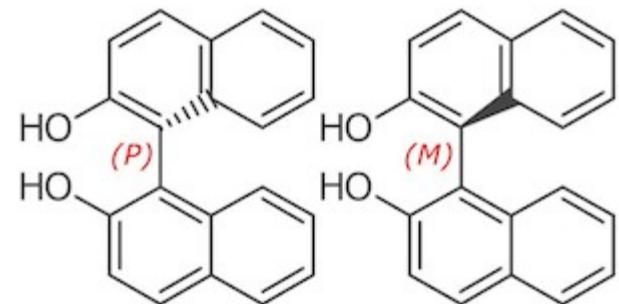
The Maven plugins were updated to allow building on Java 17



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

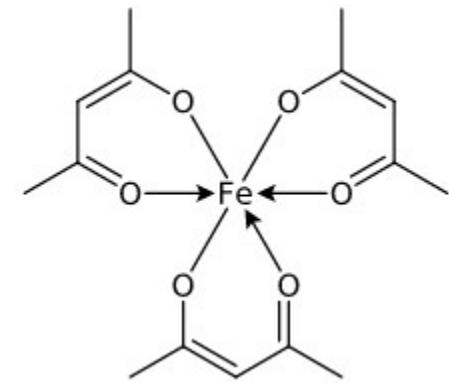
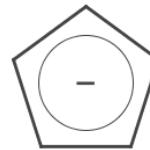
CDK 2.2 (2018-10)

- 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)

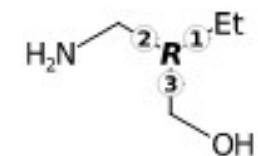
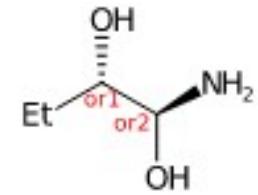
- JPLogP (Lhasa)
- fixed ALogP
- donuts are back
- other rendering options



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)

- Element 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



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>

The state of the CDK

Egon Willighagen
@egonwillighagen
0000-0001-7542-0286

2022-04-05, #cdk20y
CC-BY, doi:10.5281/zenodo.6414204