

# Further statistics on the papers citing the CDK

Egon Willighagen 

Published February 22, 2010

## Citation

Willighagen, E. (2010). Further statistics on the papers citing the CDK. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/65fss-f1y18>

## Keywords

Cdk

## Copyright

Copyright © Egon Willighagen 2010. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## chem-bla-ics

I already gave a [wordle](#) of the titles of papers citing the [first CDK paper](#). Below follows some additional statistics: the number of papers that use a particular [CDK](#) package (51). Now, these numbers are a bit rough, and surely any paper that uses the CDK is bound to use the IO or SMILES package too. Additionally, for 10 papers I was not sure what CDK functionality they used, so I assigned those to the root package.

```
org.openscience.cdk.qsar: 12 (~20%)
org.openscience.cdk: 10
org.openscience.cdk.fingerprint: 9
org.openscience.cdk.isomorphism: 6
org.openscience.cdk.similarity: 3
org.openscience.cdk.smiles: 2
org.openscience.cdk.io.cml: 2
org.openscience.cdk.model.builder3d: 2
org.openscience.cdk.ringsearch: 2
org.openscience.cdk.tools: 2
org.openscience.cdk.render: 1
org.openscience.cdk.structgen: 1
org.openscience.cdk.graph.matrix: 1
org.openscience.cdk.io: 1
```

From this we learn what parts of the CDK are used. From the various CDK Literature blogs ([#1](#), [#2](#), [#3](#), [#4](#), and [#5](#)) I already knew the descriptor calculation was much used, as well as the fingerprinter and the isomorphism checker which also provide the maximum-common substructure functionality. What I was not aware of, is that our 3D model builder had been used in published studies too, which was a pleasant surprise.

These numbers are based on 51 papers where CDK functionality was used, but you may be aware that [Web-of-Science](#) has 84 papers citing the first CDK paper. Of these, only 78 are actually in their database (which I don't quite understand). Also, at least some 10 papers cite the CDK, but do not use it, and a few papers cite the CDK where they actually use [Jmol](#). I also have to say, that for a curated citation database, I too often have to send in bug reports, but I cannot estimate to what extent that affects these numbers.

What does affect these numbers, is that some papers do not explicitly cite the CDK through one of the two papers, but only the website, or not at all (yes, that [happens](#), but it nicely balances out with papers citing the CDK but using [Jmol](#) :).

Well, I'm curious what the statistics will say about the [second CDK paper](#), and the [JChemPaint paper](#) which is based on the CDK too.