

QSAR plugin for Bioclipse getting in shape

Egon Willighagen 

Published June 27, 2007

Citation

Willighagen, E. (2007). QSAR plugin for Bioclipse getting in shape. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/ht8k2-aed07>

Keywords

Bioclipse, Qsar, Cdk, Ambit

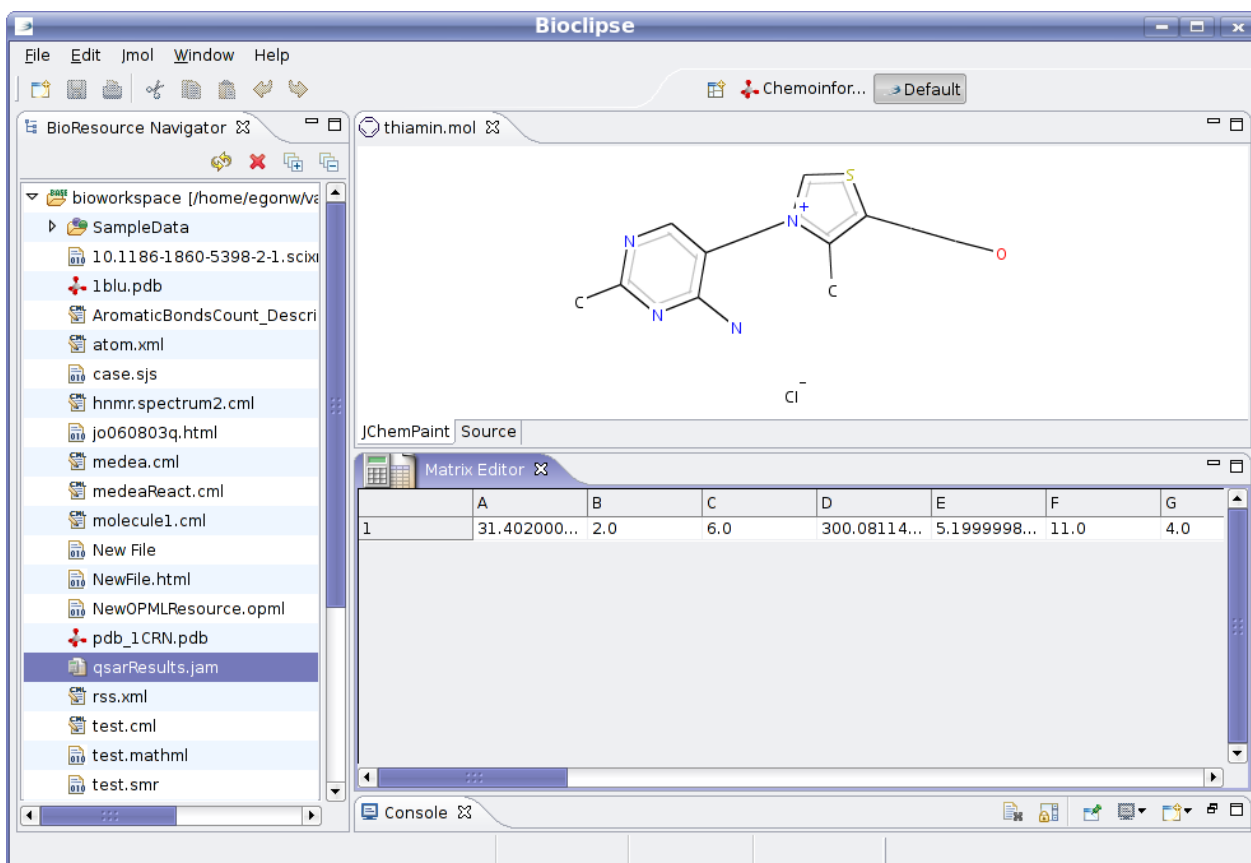
Copyright

Copyright © Egon Willighagen 2007. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

chem-bla-ics

Over the last few weeks I continued the work on getting (descriptor-based) QSAR/QSPR implemented in Bioclipse. JOELib (GPL) and the CDK (LGPL) being two prominent opensource engines that can calculate molecular descriptors, and AMBIT a front-end.

To be able to do QSAR/QSPR model building from start to end in Bioclipse, I worked in April on an architecture for selecting descriptors. Being busy with so many things, it took me some time to get around to completing that, but here are the screenshots:



The screenshot displays the Bioclipse application window. The main canvas shows the chemical structure of thiamin (thiazolopyrimidin-4(5H)-one) with a chloride ion (Cl⁻) nearby. The interface includes a menu bar (File, Edit, Jmol, Window, Help), a toolbar, and a BioResource Navigator on the left. The navigator shows a file tree with 'qsarResults.jam' selected. Below the structure is a 'JChemPaint Source' panel. At the bottom, a 'Matrix Editor' window is open, displaying a matrix with the following data:

	A	B	C	D	E	F	G
1	31.402000...	2.0	6.0	300.08114...	5.1999998...	11.0	4.0

The funny characters and the whitespace is gone. Right now, it still only lists one provider, but I plan to add JOELib plugin soon. The list of actual descriptors is provided by the extension.

What Bioclipse then does, is have the extension calculate the descriptor values for the selected CDKResource in the BioNavigator using the selected descriptors. This will then create a new MatrixResource in the Bioclipse workspace (currently called qsarResult.jam), and which is opened in the Matrix editor:

The screenshot displays the Bioclipse application window. The main workspace shows the chemical structure of thiamin (vitamin B1) with a chloride ion (Cl⁻) below it. The structure consists of a pyrimidine ring connected to a thiazolium ring, which has a methyl group and a hydroxyl group. The label 'thiamin.mol' is visible in the top right of the workspace.

On the left, the BioResource Navigator shows a file tree under 'bioworkspace [home/egonw/vz]'. The 'SampleData' folder is expanded, listing various files including '10.1186-1860-5398-2-1.scixf', '1blu.pdb', 'AromaticBondsCount_Descri', 'atom.xml', 'case.sjs', 'hnmr.spectrum2.cml', 'jo060803q.html', 'medea.cml', 'medeaReact.cml', 'molecule1.cml', 'New File', 'NewFile.html', 'NewOPMLResource.opml', 'pdb_1CRN.pdb', 'qsarResults.jam', 'rss.xml', 'test.cml', 'test.mathml', and 'test.smr'. The 'qsarResults.jam' file is currently selected.

At the bottom, the Matrix Editor window is open, showing a single row of data:

	A	B	C	D	E	F	G
1	31.402000...	2.0	6.0	300.08114...	5.1999998...	11.0	4.0

The Matrix Editor window also has a 'Console' tab at the bottom.

There is still enough work left to do. For example, the columns are not yet labeled according to the descriptor name, and selecting more than one `CDKResource` in the navigator does not give a multirow matrix yet.