

New Bioclipse Features: Kabsch Alignment, RMSD Distance and Tanimoto Similarity Matrices

Egon Willighagen 

Published November 4, 2009

Citation

Willighagen, E. (2009). New Bioclipse Features: Kabsch Alignment, RMSD Distance and Tanimoto Similarity Matrices. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/2ty9f-n0m97>

Keywords

Cheminf, Cdk, Bioclipse, Jmol

Copyright

Copyright © Egon Willighagen 2009. 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

We recently submitted a second paper on [Bioclipse](#), and have worked hard in the past two weeks on addressing the reviewers' questions (and we love these feature requests! See also these [two blogs](#)). One reviewer seemed very interested in seeing docking available in Bioclipse. While we do not have a full docking feature set up for Bioclipse, we do have functionality to deal with 3D structures, though our research urged us to focus on the 2D side of cheminformatics so far.

To strengthen our intentions towards the 3D cheminformatics world, we have implemented a few new features, using [CDK](#) functionality. For example, we added Kabsch alignment and the related RMSD between molecular structures implemented as both popup menus as well as manager methods. The manager method you can see in action in [MyExperiment workflow 937](#), which you can download directly into Bioclipse with one simple command (see [Bioclipse Manager for MyExperiment.org](#)):

```
var smileses = new Array("CC(C)C", "CCCN", "CCC=O");

var unaligned = cdk.createMoleculeList();
for (i=0; i<smileses.length; i++) {
    mol = cdk.fromSMILES(smileses[i]);
    mol = cdk.generate3dCoordinates(mol)
    unaligned.add(mol);
}

var aligned = cdk.kabsch(unaligned)

jmol.load(aligned.get(0));
for (i=1; i<aligned.size(); i++) {
    jmol.append(aligned.get(i));
}
```

Now, we do have to update the use of Jmol in Bioclipse, and a big overhaul is scheduled for the 2.4 released in February next year. But you get the idea.

As said, there are two stories to adding this new functionality. Because we want all GUI interaction the user performs to be recordable (Scientist 1: *What did you do to get those nice results?* Scientist 2: *I pushed that button in the that long menu.* Scientist 1: *What button is that?* Scientist 2: *Wait, I send you the BSL script with a Google Wave.*)

The managers that allow this recording is Bioclipse specific, and also the reason why it would not be trivial to make a general Bioclipse plugin for Eclipse... some Spring magic is used to inject the managers into the JavaScript language. Anyway, the second thing is to add a GUI element, like popup menus. Now, this is a particular area where Eclipse excels. Now, I did have to ask for the details, as I am not using this daily (I'm doing science, not IT), but Ola was kind enough to give me the pointers for it.

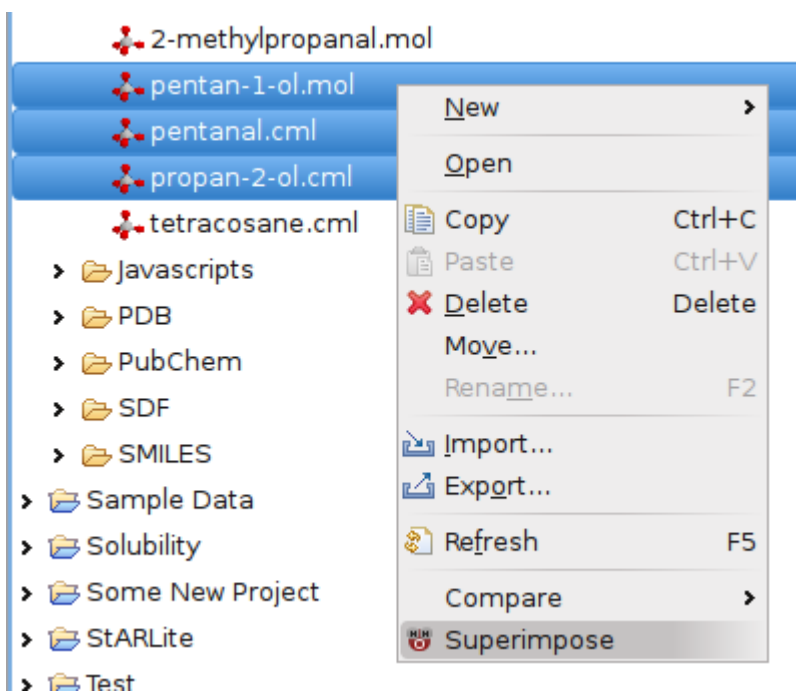
chem-bla-ics

The below configuration snippet links the pop up action to Bioclipse Navigator content (you know, where your MDL SD, CML, script and other files show up in Bioclipse). **But** only if I have selected 3 or more files! And, only if those files are actually some molecular content with 3D coordinates! And Bioclipse inherits this functionality by using the Eclipse platform.

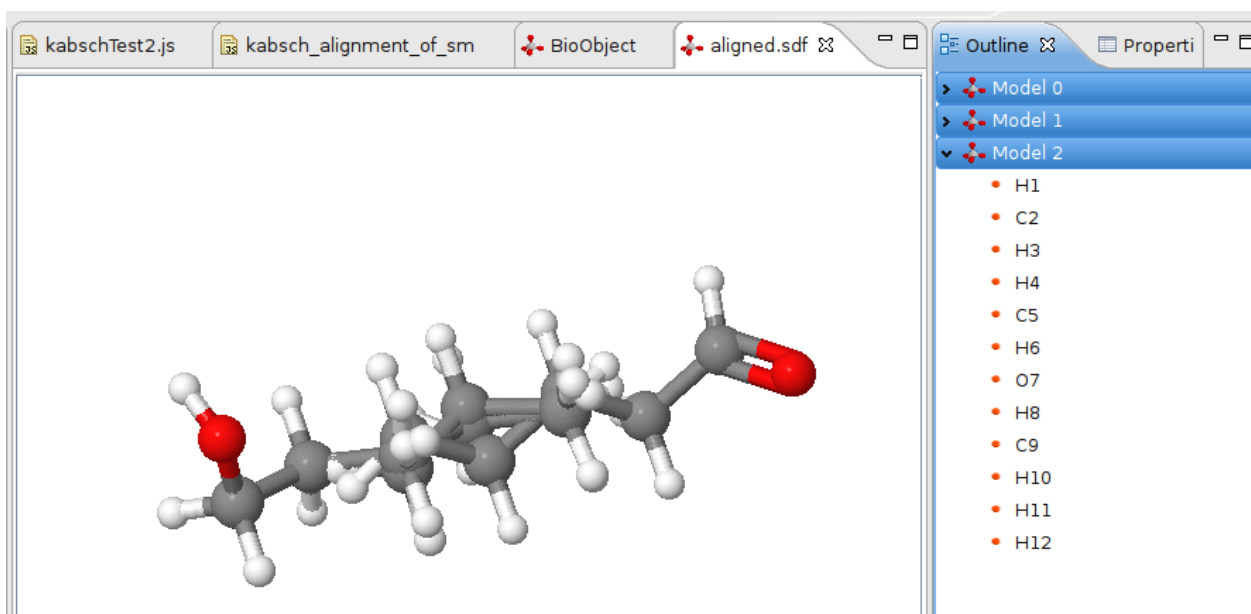
```
<menuContribution
  locationURI="popup:org.eclipse.ui.popup.any?after=additions">
  <command
    commandId="net.bioclipse.cdk.ui.handlers.kabschAlignment"
    label="Perform Kabsch Alignment"
    icon="icons/molecule2D.png">
  <visibleWhen>
    <with variable="selection">
      <count value="(2-"/>
      <iterate operator="and" ifEmpty="false">
        <adapt type="org.eclipse.core.resources.IResource">
          <or>
            <test property="org.eclipse.core.resources.contentType"
              value="net.bioclipse.contenttypes.cml.singleMolecule3d"/>
            <test property="org.eclipse.core.resources.contentType"
              value="net.bioclipse.contenttypes.cml.singleMolecule5d"/>
            <test property="org.eclipse.core.resources.contentType"
              value="net.bioclipse.contenttypes.mdlMolFile3D"/>
          </or>
        </adapt>
      </iterate>
    </with>
  </visibleWhen>
</command>
</menuContribution>
```

When Bioclipse is run, this looks like:

chem-bla-ics

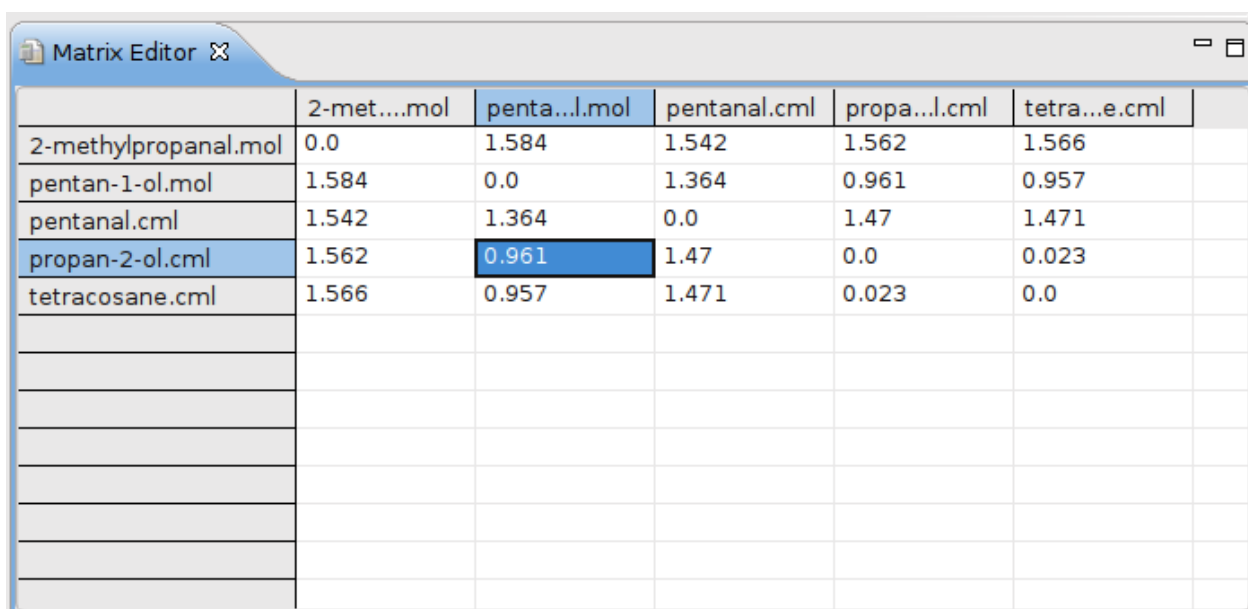


And the alignment results will nicely show up in a Jmol viewer (while it is implemented as an Eclipse editor, it is not yet):



The first screenshot also shows the new pop-up menus for calculating two matrices for 3 or more molecules. One is based on the [RMSD](#) of the 3D atomic coordinates of the atoms in the [MCSS](#) (BTW, Asad's SSSD work is making its way into the CDK library, and will be available in a later Bioclipse version too.) and will create a distance matrix. The second new pop-up menu used the Tanimoto similarity measure based on CDK fingerprints on the selected chemical graphs. If the Bioclipse Statistics feature is installed, the created [CSV](#) files will open up in a matrix editor:

chem-bla-ics



	2-met....mol	penta...l.mol	pentanal.cml	propa...l.cml	tetra...e.cml
2-methylpropanal.mol	0.0	1.584	1.542	1.562	1.566
pentan-1-ol.mol	1.584	0.0	1.364	0.961	0.957
pentanal.cml	1.542	1.364	0.0	1.47	1.471
propan-2-ol.cml	1.562	0.961	1.47	0.0	0.023
tetracosane.cml	1.566	0.957	1.471	0.023	0.0

Kabsch alignment of protein backbones is planned for a later Bioclipse release, but an important feature for [our groups proteochemometrics work](#).