

CIP rules for stereochemistry

Egon Willighagen 

Published April 22, 2010

Citation

Willighagen, E. (2010). CIP rules for stereochemistry. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/1n6nz-z2572>

Keywords

Cdk, Chemistry, Cheminfo, Scholia

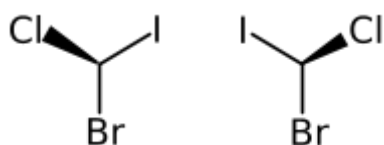
Copyright

Copyright © Egon Willighagen 2010. 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

Uniquely identifying [stereochemical enantiomers](#) is an important aspect of data exchange of chemical structures. The simplest, most neglected solution is to pass around 3D models, but a lot of people like to stick to things like [SMILES](#), or [IUPAC names](#). Now, given that we want to uniquely represent the stereochemistry, we can use special rules. One example for enantiomers are the [Cahn-Ingold-Prelog \(CIP\) rules](#).

The [CDK](#) does not have an implementation of (part of) the CIP rules. However, we recently started a collaboration with Dr [Lars Carlsson](#) in the Computational Toxicology, Global Safety Assessment group at [AstraZeneca R&D Mölndal](#), headed by Dr Scott Boyer. Within this collaboration I have started a partial implementation of the CIP rules. The full set of rules is quite extensive, and some subrules are outside the scope of the collaboration. For example, we will likely not look at axial or helical stereochemistry within this collaboration. The kind of things it is able to do is distinguish between these mirror images (yeah, I should use [Jmol](#), but [Scholia](#) needs more plugging right now: click the images):



The current patch is not looking into the problem of which atom is chiral; that problem is quite complex in itself, and Tim is writing up a nice set of [blogs about that](#). Further, the current aims focuses only at application to atoms of ligancy four; that is, carbons.

The CIP rules uniquely define the stereochemistry of such a carbon, by uniquely ordering the ligands around the atom. Using rules the ligands are ordered, and they include rules defining priority based on atomic number, mass number, etc. It is the recursion that makes things more interesting, but I will not delve into the details of the algorithm here (see the [aforelinked Wikipedia page](#) instead, or a [cheminformatics book](#) like the one shown on the right). Here, I want to introduce some of the API of the current patch for the CDK.

Ligands and their Priorities

Core to the implementation are the CIP priority rules, that allow ordering of the ligand. So, we define a molecule, and ligands:

```
IMolecule molecule = parser.parseSmiles("IC(Br)(Cl)[H]");
ILigand ligand1 = new Ligand(
    molecule.getAtom(1), molecule.getAtom(2)
);
ILigand ligand2 = new Ligand(
    molecule, molecule.getAtom(1), molecule.getAtom(0)
);
ISequenceSubRule rule = new CIPLigandRule();
```

chem-bla-ics

```
Assert.assertEquals(-1, rule.compare(ligand1, ligand2));  
Assert.assertEquals(1, rule.compare(ligand2, ligand1));
```

This [JUnit](#) test looks at the chiral compound given earlier, but without specifying the stereochemistry using the @@/@ SMILES syntax; we get to that later. Here, the example defines two ligands around atom 1 (which is the carbon; the index starts at 0). The first ligand is the bromine, the second ligand is the iodine. Because the latter takes priority according to the CIP rules, the `compare(ligand1, ligand2)` returns -1.

The CIPTool

This *CIPLigandRule* is used in the CIPTool to provide more user-oriented methods. The goal, obviously, is this bit of code:

```
IMolecule molecule = parser.parseSmiles("ClC(Br)(I)[H]");  
LigancyFourChirality chirality =  
    CIPTool.defineLigancyFourChirality(  
        molecule, 1, 4, 0, 2, 3, STEREO.CLOCK_WISE  
    );  
Assert.assertEquals(  
    CIP_CHIRALITY.R,  
    CIPTool.getCIPChirality(chirality)  
);
```

Because we do not have 3D coordinates in our SMILES, we define the stereochemistry as `CLOCK_WISE` and `ANTI_CLOCK_WISE`. The former here means that, looking from the first ligand, following atoms 2, 3, and 4 are oriented in a circle in a clock-wise turn. This defines uniquely the geometrical orientation, but which changes between `CLOCK_WISE` and `ANTI_CLOCK_WISE` upon every atom-atom exchange. Therefore, we uniquely prioritize the ligands, project, and translate the resulting `CLOCK_WISE` or `ANTI_CLOCK_WISE` in the appropriate R and S stereochemistry.

That's all for now. Questions, ideas and others most welcome in the comment!