

CIP rules #2: parsing @ and @@ from SMILES

Egon Willighagen 

Published May 8, 2010

Citation

Willighagen, E. (n.d.). In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/n8f2y-adq94>

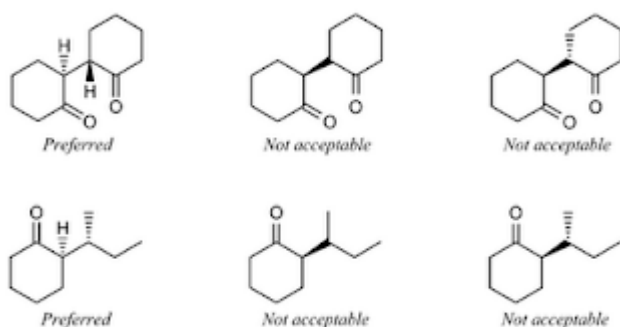
Keywords

Cdk, Iupac

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.

I [recently wrote](#) about a project for a (partial) CIP implementation. This implementation is in place, and we are working towards setting up an extensive test suite. The data set we had in mind was available as SMILES and as MDL molfile. Now, the latter does not really specify the stereochemistry of the tetrahedral centers, and relies on wedge bonding. Actually, a few years ago Jonathan Brecher wrote up the IUPAC recommendation for the use of the wedge bond for chirality specification (doi:[10.1351/pac200678101897](https://doi.org/10.1351/pac200678101897)), with 74 pages of rules and examples, like the following (copyright by authors or journal; I'm claiming fair use):



So, using wedges leaves plenty of room for incorrectly specifying the stereochemistry. Therefore, we decided to go for SMILES, even though Noel [recently showed](#) that processing stereochemistry in SMILES is not trivial either. The SMILES I am currently using:

- Br[C@@H](Cl)I
- Br[C@H](Cl)I
- Br[C@@]([H])(Cl)I
- Br[C@]([H])(Cl)I
- [C@]12(OC1)NCN2
- C[C@H](O)[C@H](O)C
- NC([C@H](O)C)Cl
- I1.Cl2.Br3.[C@]123CCC

I'm looking for more corner cases... please leave them as comments.