

# SMILES to become an Open Standard

Egon Willighagen 

Published September 28, 2007

## Citation

Willighagen, E. (2007). SMILES to become an Open Standard. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/8c47c-8dq17>

## Keywords

Smiles, Blue-obelisk

## Copyright

Copyright © Egon Willighagen 2007. 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

Craig James wants [to make SMILES an open standard](#), and this has been received with much enthusiasm. SMILES ([Simplified molecular input line entry specification](#)) is a de facto standard in cheminformatics, but the specification is not overly clear, which Craig wants to address. The [draft](#) is CC-licensed and will be discussed on the new [Blue Obelisk blueobelisk-smiles](#) mailing list.

Illustrative is my confusion about the sp<sup>2</sup> hybridized atoms, which use lower case element symbols in SMILES. Very often this is seen as indicating aromaticity. I have written up [the arguments supporting both views](#) in the [CDK wiki](#). I held the position that lower case elements indicated sp<sup>2</sup> hybridization, and the CDK SMILES parser was converted accordingly some years ago. A recent discussion, however, stirred up the discussion once more (which led to the aforementioned wiki page).

You can imagine my excitement when I looked up the meaning in the new draft. It states: *The formal meaning of a lowercase "aromatic" element in a SMILES string is that the atom is in the sp<sup>2</sup> electronic state. When generating a normalized SMILES, all sp<sup>2</sup> atoms are written using a lowercase first character of the atomic symbol. When parsing a SMILES, a parser must note the sp<sup>2</sup> designation of each atom on input, then when the parsing is complete, the SMILES software must verify that electrons can be assigned without violating the valence rules, consistent with the sp<sup>2</sup> markings, the specified or implied hydrogens, external bonds, and charges on the atoms.*