

Evidence of Aromaticity

Egon Willighagen 

Published November 6, 2007

Citation

Willighagen, E. (2007). Evidence of Aromaticity. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/9fkkg-fxz59>

Keywords

Cdk, Aromaticity, Crystal

Abstract

I have been working on a new atom type perception engine for the CDK, after having decided that the existing atom type lists were not sufficient for the algorithms we have in the CDK.

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.

I have been working on a new atom type perception engine for the [CDK](#), after having decided that [the existing atom type lists](#) were not sufficient for the algorithms we have in the CDK. The [new list](#) is growing in size, and basically contains four properties (besides element and formal charge):

1. number of bounded neighbors
2. number of pi bonds (or double bond equivalents)
3. number of lone pairs
4. hybridization state

This seems to be a minimal and accurate set to cover a rather good deal of chemoinformatics. I have yet to make the mappings of the new atom type list with existing lists for force fields, and radicals are missing too. However, the following algorithms in the CDK seem to translate rather well:

- hydrogen adding
- aromaticity detection (Hückel rules)

I still have to rework the double bond perception.

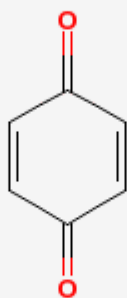
Aromaticity

Now, aromaticity is a fuzzy concept, and there is no general agreement on what it is. Some say it is smelly compounds, others say ring systems which apply to the Hückel rule. Based on the new atom type list, I have rewritten the Hückel aromaticity detector and it applies these rules:

- only single rings and two fused non-spiro rings
- $4n+2$ electrons
- no ring atoms with double points not in the ring too

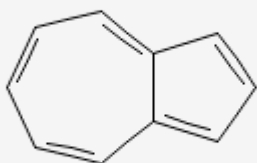
This approach differs in two ways from the old code: it no longer tries to test all ring systems, which required to use the [CDK AllRingsFinder algorithm](#) which combinatorial generates all possible ring systems. The new code only considers ring systems with up to two single rings. Aromaticity beyond that is even less well defined than aromaticity in general.

The other difference is that the ring system must not have ring atoms which have a double bond which is not part of the ring too. The classical example is benzoquinone (InChI=1/C6H4O2/c7-5-1-2-6(8)4-3-5/h1-4H) which is not aromatic, even though it conforms the $4n+2$ rule (image from [PubChem](#)):



Evidence of Aromaticity

The final rule, of course, is what nature tells us what is aromatic and what is not. There are many other details to aromaticity than I just covered. For example, take azulene (InChI=1/C10H8/c1-2-5-9-7-4-8-10(9)6-3-1/h1-8H). All atoms are aromatic, but not all bonds (also [PubChem](#)):



These things are complex, but the rise of [Open Data](#) helps us out, as well as increasing computing power. [Peter](#) has been running two rather projects which may help us out: [CrystalEye](#) (Nick: no blog?) and [OpenNMR](#).

NMR shifts will give us experimental backup on our notion of aromaticity, and so do bond lengths. I [asked Peter about this](#), and whether OpenNMR predicted shifts could indeed confirm aromaticity of compounds, and [he replied](#) and showed that the predicted spectra could be used to distinguish between C-C and C=C bonds.

I commented the following (which was in moderation at the time of writing), and that gets us to experimental evidence for aromaticity:

Thanx for the elaborate answer. What I had in mind was the question whether NMR shift predictions can be used to tell me if a certain ring system is aromatic or not, and in case of fused rings, which atoms and which bonds are aromatic and which not. I'm sure the prediction error for ^1H NMR shifts is well below 2ppm, and more in the order of 0.2ppm.

But maybe I should be asking, can I use CrystalEye to decide if ring systems are "aromatic", and in case of two rings fused together (non-spiro), which atoms and bonds are aromatic and which not. Aromaticity is a fuzzy concept, with various definitions. I would be interesting in linking what the expert considers 'aromatic' (or SMILES, or the CDK, or ...) with what the QM chemistry (via bond lengths or NMR shift predictions) and crystal structures (via bond lengths) has to teach us. The null hypothesis being that the bonds are not delocalized (bond length) and that no ring current is found (NMR shifts, ^1H in particular).

Regarding those bond lengths, 'aromatic' bonds show a bond length in between that of single and double bonds (e.g. see [this random pick](#)). The CrystalEye data does not reflect that really, and only [a trimodal histograms](#) shows up. Indeed, the $\text{C}\equiv\text{C}$ peak is very low, around 1.2Å :) Apparently, the triple $\text{C}\equiv\text{C}$ bond order is underrepresented in nowadays crystallography.

Maybe aromatic $\text{C}:\text{C}$ bonds are underrepresented too, or can the absence of a peak around 1.40Å be explained otherwise? I would at least have expected a shoulder or deviation in peak shape of the peak at 1.37Å.

This is what the histogram looks like (for archival reasons):

C-C Bond Lengths in CrystalEye (Last updated 2007-11-06)
for non-disordered, unconstrained atoms in structures where temperature ≤ 200.0 and r-factor ≤ 0.05

