Counting constitutional isomers from the molecular formula

Egon Willighagen ⬤

Citation

Keywords

Copyright

chem-bla-ics

Update: check these two papers.

We all know the combinatorial explosion when calculating the number of possible constitutional isomers (see wp:structural isomorphism) of a certain molecular formula. For example, C2H6 has only one constitutional isomer (ethane, [InChI=1/C2H6/c1-2/h1-2H3] {.chem:inchi xmlns:chem="http://www.blueobelisk.org/chemistryblogs/"}), and C4H10 has only two. Especially, breaking symmetry by replacing one carbon by another element, or replacing a single by a double bond, increases the number sharply. For example, C7H16 has only nine constitutional isomers, while replacing two single bonds by two double bonds, creating C7H10, increases this number to 499! Then, replacing in the last formula, one carbon by an oxygen adds another few, totaling 747 isomers.

Now, C8H8NBr has at least **649 thousand** constitutional isomers, and I am quite interested in being able to know the number of isomers beforehand, without having to generate the structures itself (for example, using CDK's `GENMDeterministicGenerator`). [InChI=1/C8H8BrN/ c9-7-1-2-8-6(5-7)3-4-10-8/h1-2,5,10H,3-4H2]{.chem:inchi xmlns:chem="http://www.blueobelisk.org/ chemistryblogs/"} is one of the isomers.

So, my question: is anyone aware of free code (in order of preference: 1. LGPL, 2. BSD/MIT,

1. opensource, 4. free) to calculate or estimate the number of constitutional isomers for a certain molecular formula. An estimate would already be nice. Ideally, I would implement this bit of code into the CDK, but otherwise, just knowing the number of isomers for C8H8NBr would be nice :)

Additionally, any relevant, recent literature recommendations are most welcomed. I am aware of the use of polynomials, but literature I have seen so far just focuses on molecules of a certain architecture, and it not able to come up with a guess based on the molecular formula alone.