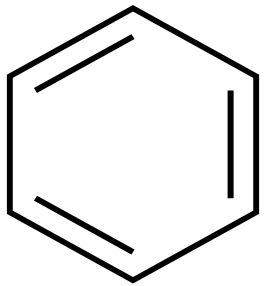
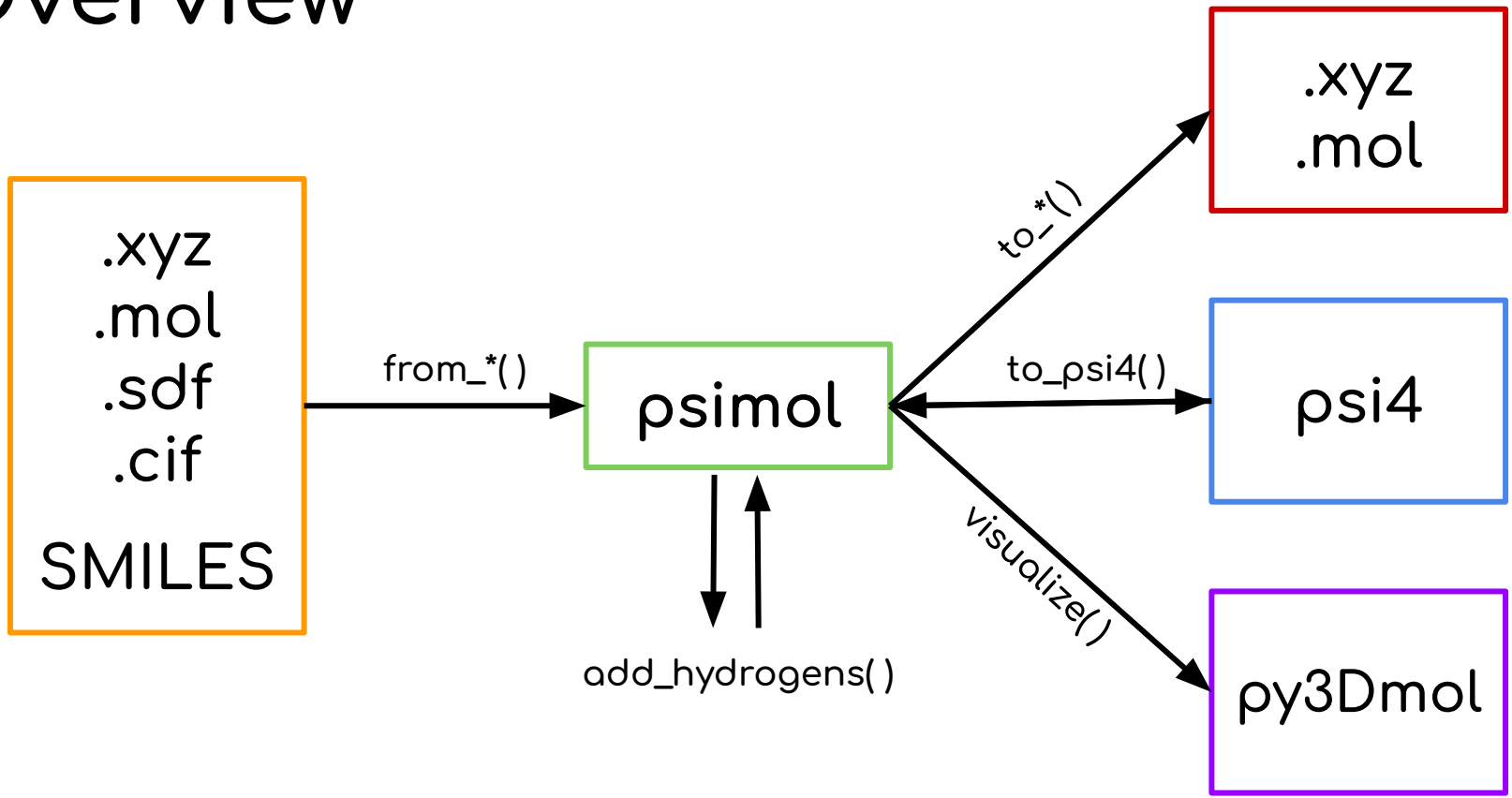
PsiMol 

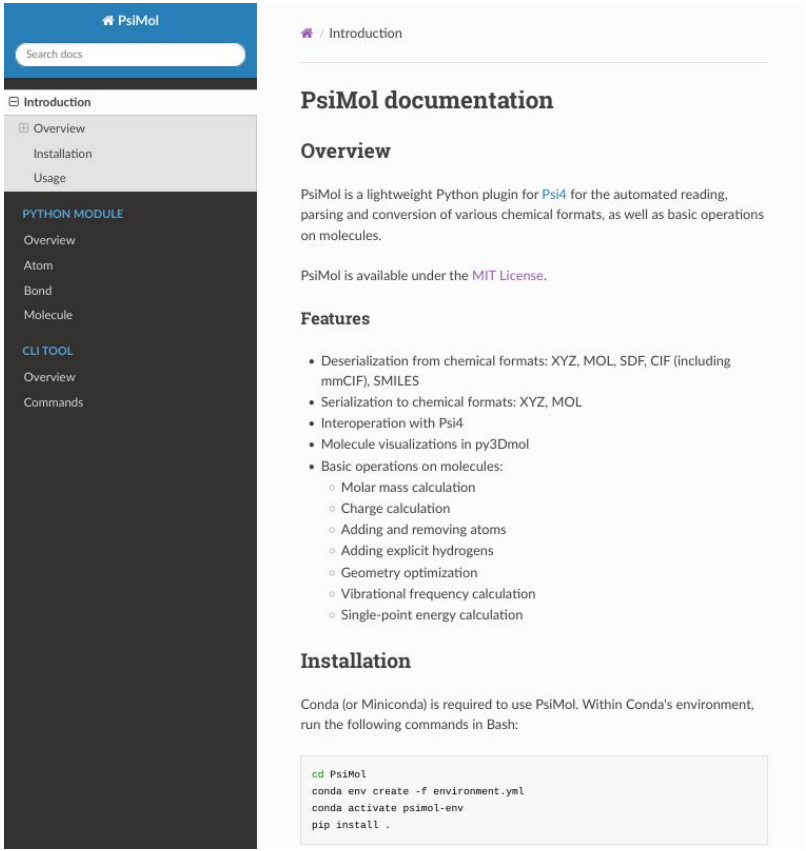
Dawid Uchal, Maciej Bielecki, Paweł Nagórko
ADP 2024

Overview


.xyz
.mol
.sdf
.cif
SMILES
from_*( )
psimol
add_hydrogens( )
to_*( )
to_psi4( )
visualize( )
.xyz
.mol
psi4
py3Dmol

Project postulates

- Harmonized, annotated code
- Thorough documentation
- Sensible API
- Available as Python module and CLI tool
- Fully open-source
- OS independent



Parsing molecular files

Deserialization from:

- XYZ
- MOL
- SDF
- CIF
- mmCIF
- SMILES
- Planned support for PDB

Serialization to:

- XYZ
- MOL

Supports atoms up to 86 atomic number (Rn), excluding f-block.

Checking aromacity

- Most coordinate-based molecular files (xyz, cif) do not contain any explicit information about bonding and its types.

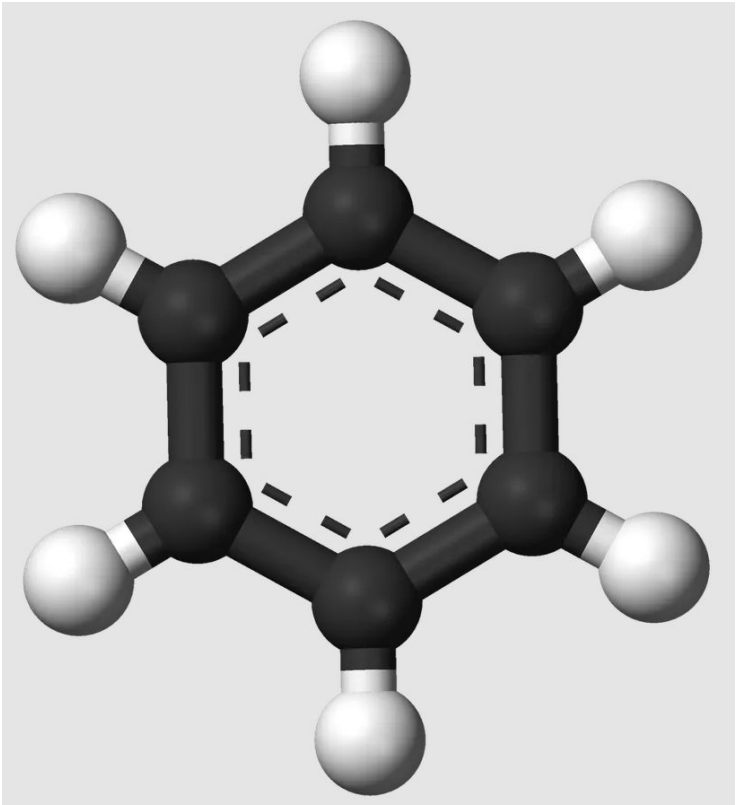
- Definition:
  - Aromatic bonds are bonds between atoms of a planar cycle consisting solely of C, N, O, S atoms

- It's an interesting algorithmic problem

Naive solution

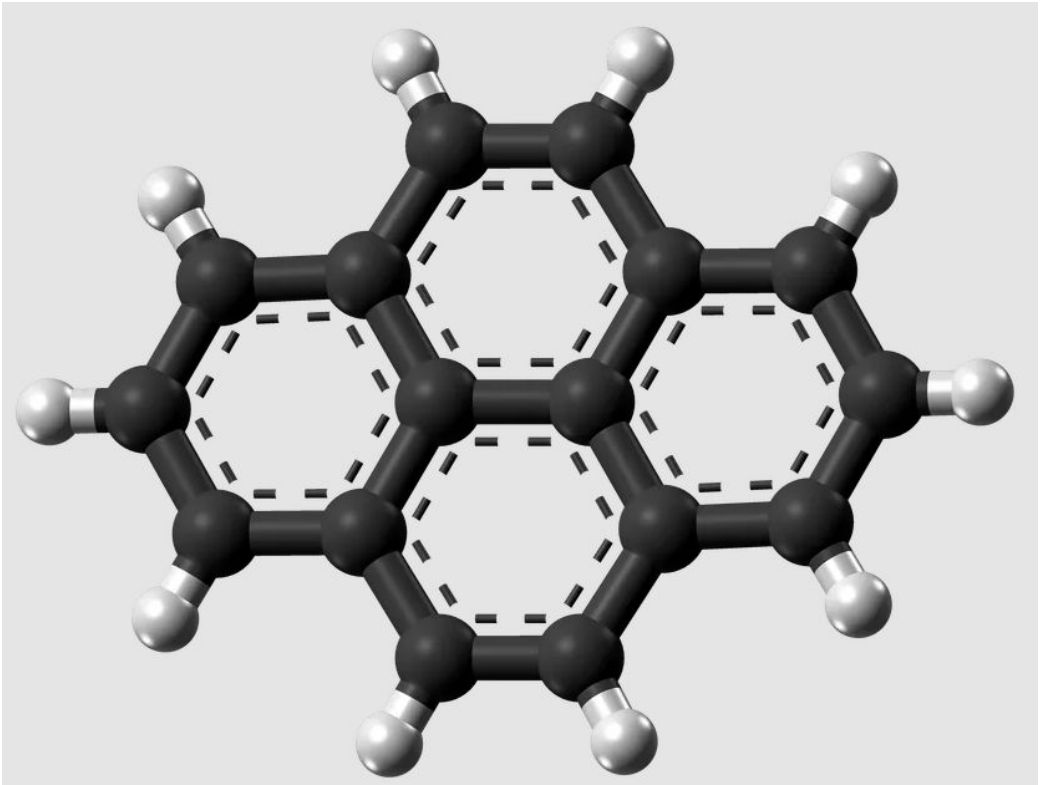
- For each atom:
  - DFS( )

- Result:
  - 6 cycles for benzene



Improvement

- Just mark atoms as visited
- Problem for pirene
- Naive solution is even worse here than expected, resulting in 280 cycles



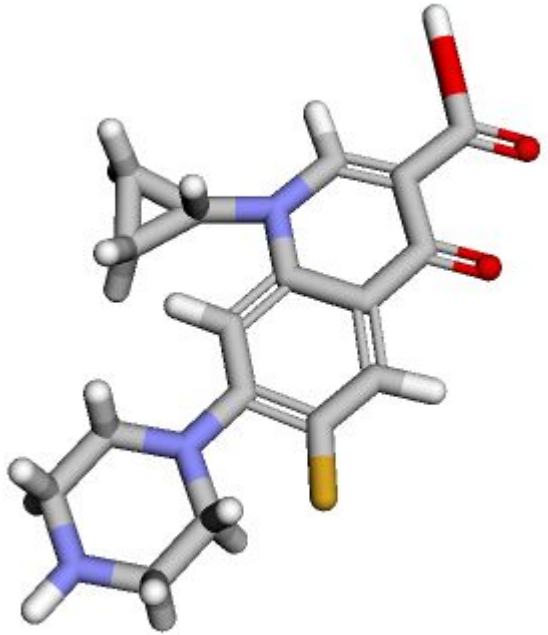
Johnson's Algorithm

- Complexity O(n+m) where n, m are vertices and edges

- Algorithm is based on coloring visited vertices with one color and with second one ones visited completely

Parsing SMILES

- PsiMol enables creating 3D molecules from their SMILES representation
- It supports *organic subset* of SMILES
- Works with most of organic compounds
- Has limitations for irregular multi-ring and non-planar clusters



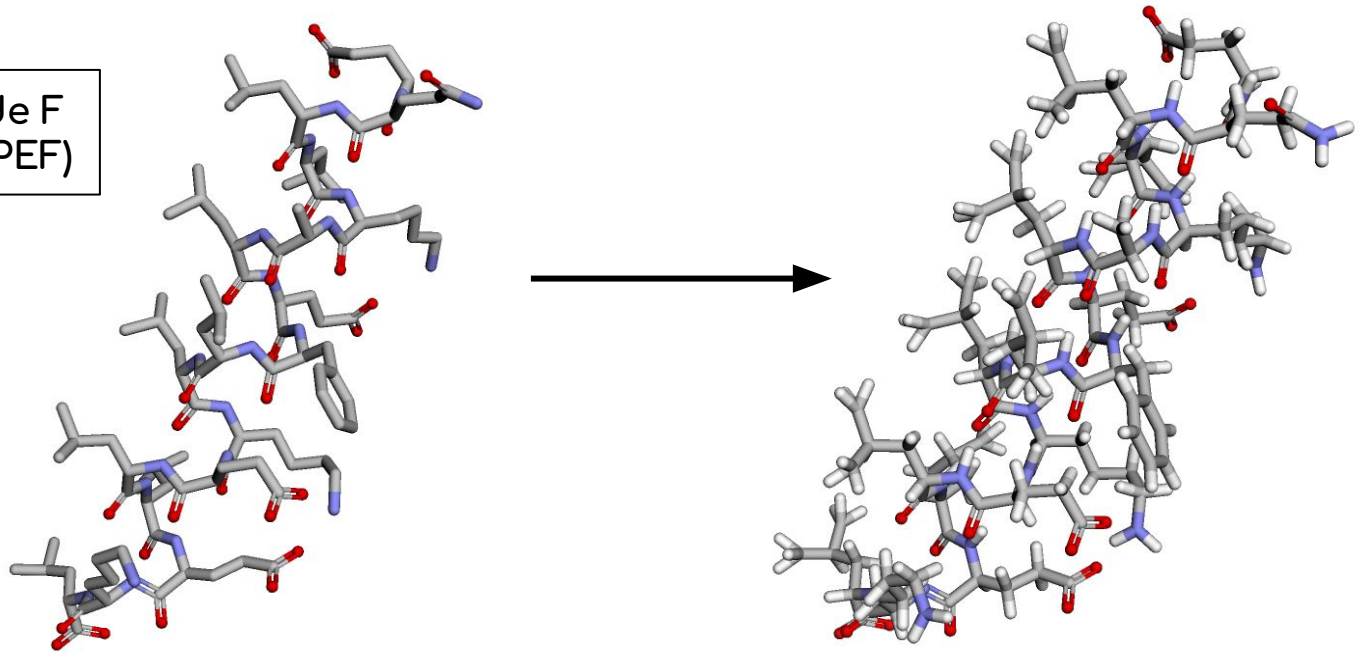
Ciprofloxacin structure made from:

C1CC1N2C=C(C(=O)C3=CC(=C(C=C32)N4CCNCC4)F)C(=O)O

Adding hydrogens

- PsiMol enables adding explicit hydrogen atoms to molecules

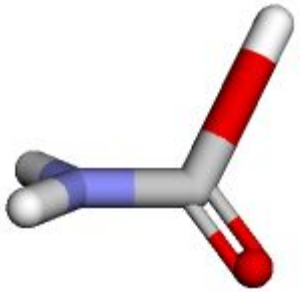
Peptide F
(PDB 1PEF)



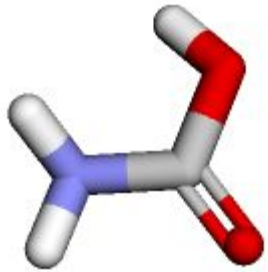
Integration with psi4

- Besides providing direct conversion to psi4.Molecule objects, PsiMol also wraps all its main utilities into own methods.

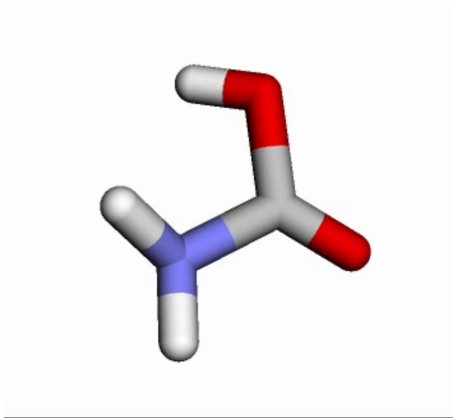
```
.from_smiles('NC(=O)O')
```



```
.optimize()
```



```
.calculate_frequencies()
.show_modes(mode=0)
```



Command line interface

- Added possibility of using our package as standalone tool
- Current commands include converting and protonating molecules

```
usage: psimol convert [-h] --input-format {xyz,cif,smiles,mol} --output-format {xyz,mol} -i INPUT
-o OUTPUT [--add-hydrogens]

options:
  -h, --help            show this help message and exit
  --input-format {xyz,cif,smiles,mol}
                        Input file format.
  --output-format {xyz,mol}
                        Output file format.
  -i INPUT, --input INPUT
                        Path to the input file.
  -o OUTPUT, --output OUTPUT
                        Path to the output file.
  --add-hydrogens       Add hydrogens to the molecule.
```

Thank you for your attention