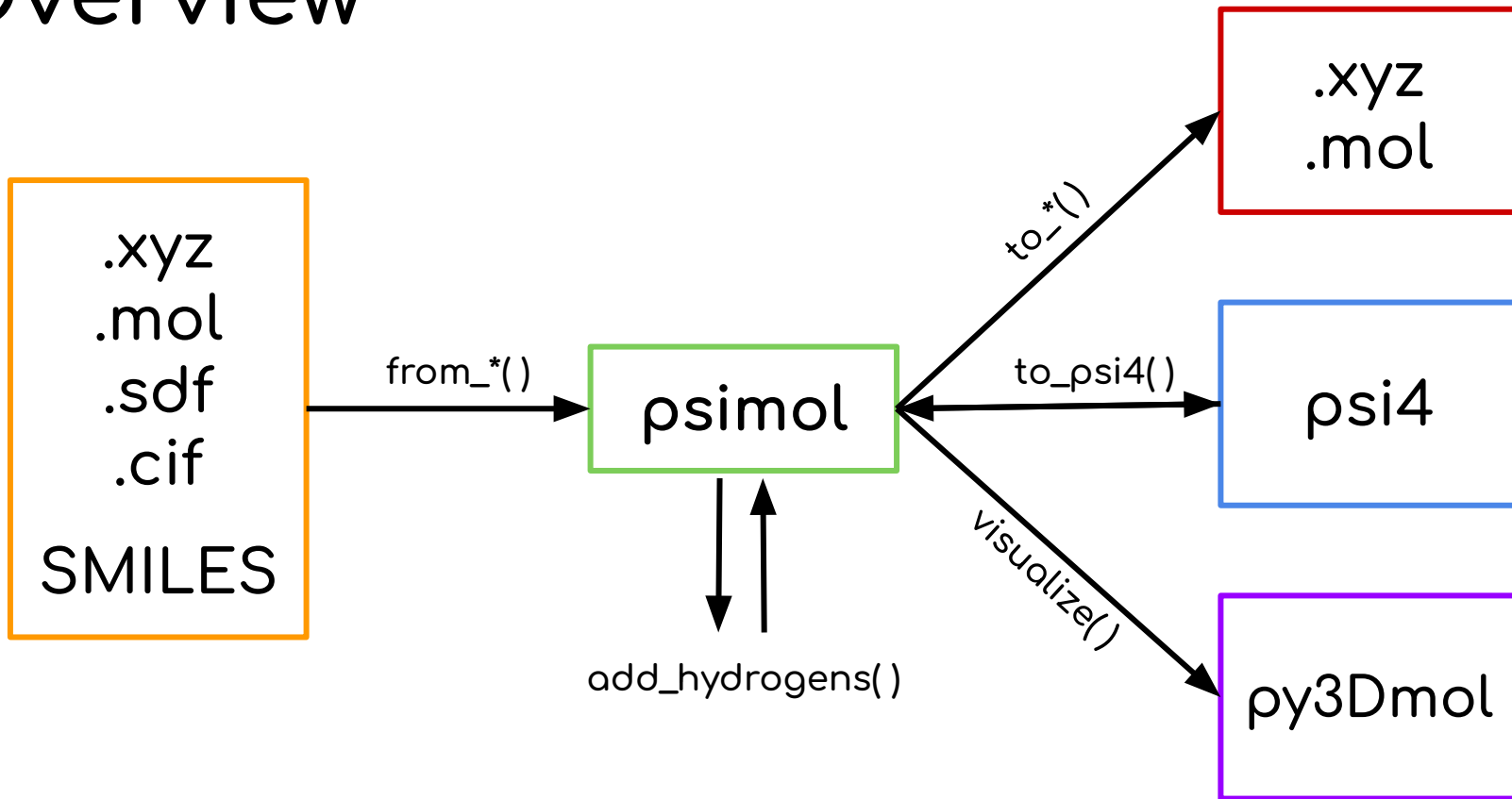


PsiMol

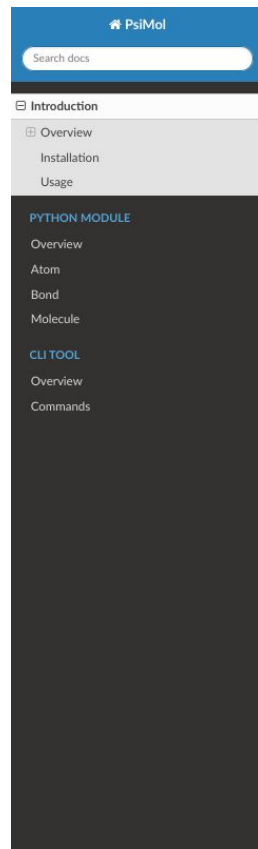
Dawid Uchal, Maciej Bielecki, Paweł Nagórko
ADP 2024

Overview



Project postulates

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



🏠 / Introduction

PsiMol documentation

Overview

PsiMol is a lightweight Python plugin for [Psi4](#) for the automated reading, parsing and conversion of various chemical formats, as well as basic operations on molecules.

PsiMol is available under the [MIT License](#).

Features

- Deserialization from chemical formats: XYZ, MOL, SDF, CIF (including mmCIF), SMILES
- Serialization to chemical formats: XYZ, MOL
- Interoperation with Psi4
- Molecule visualizations in py3Dmol
- Basic operations on molecules:
 - Molar mass calculation
 - Charge calculation
 - Adding and removing atoms
 - Adding explicit hydrogens
 - Geometry optimization
 - Vibrational frequency calculation
 - Single-point energy calculation

Installation

Conda (or Miniconda) is required to use PsiMol. Within Conda's environment, run the following commands in Bash:

```
cd PsiMol
conda env create -f environment.yml
conda activate psimol-env
pip install .
```

Parsing molecular files

Deserialization from:

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

Serialization to:

- XYZ
- MOL

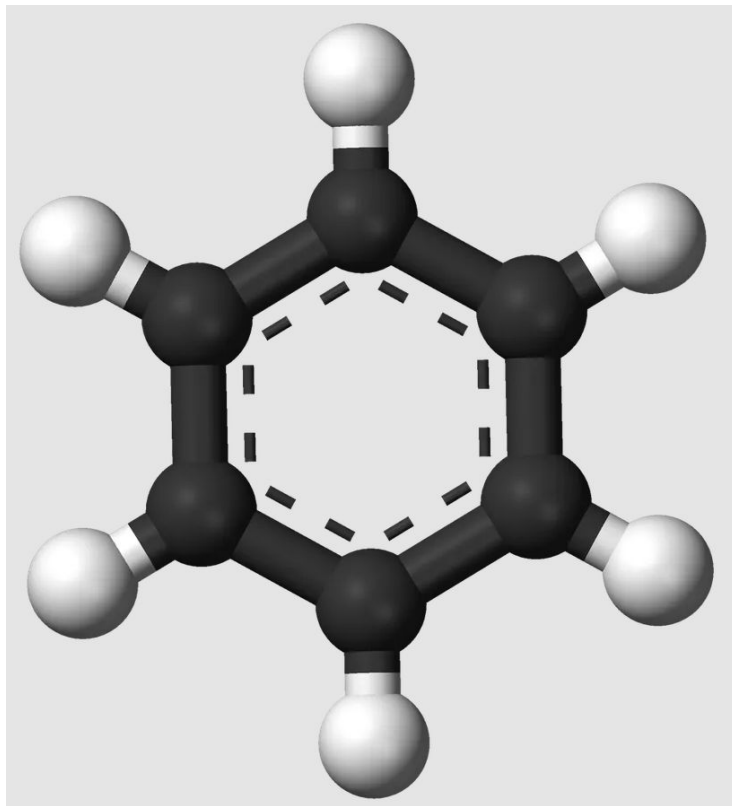
Supports atoms up to 86 atomic number (Rn), excluding f-block.

Checking aromaticity

- 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

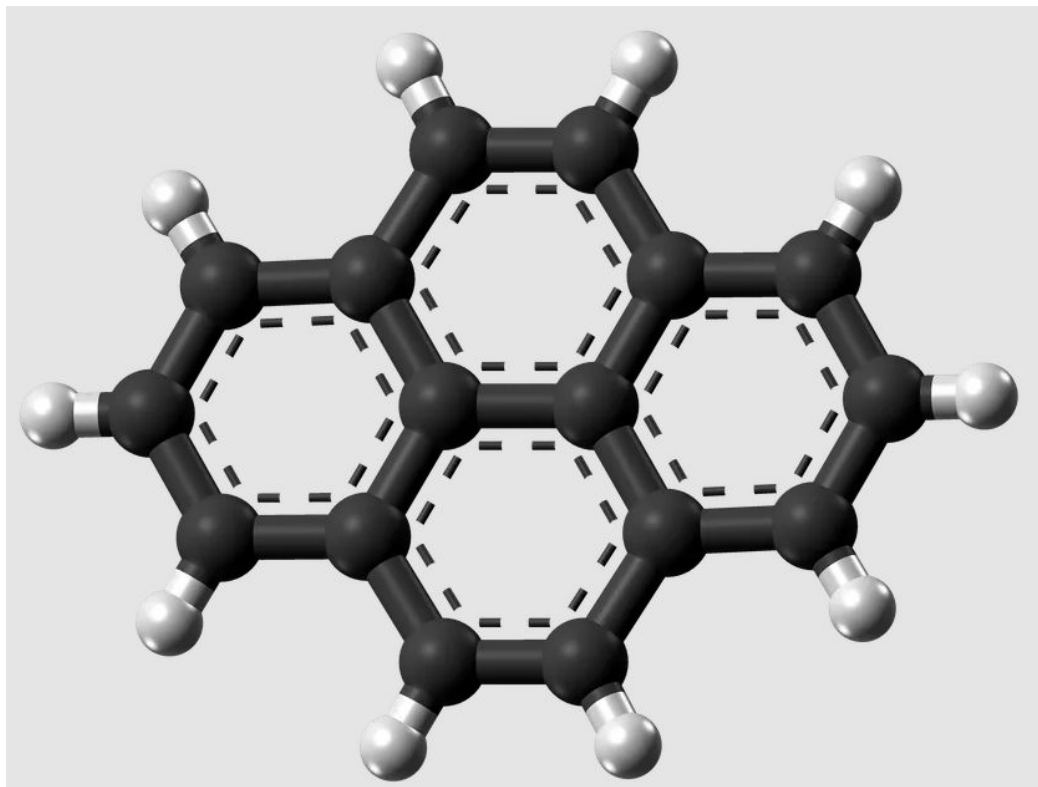
Naive solution

- 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

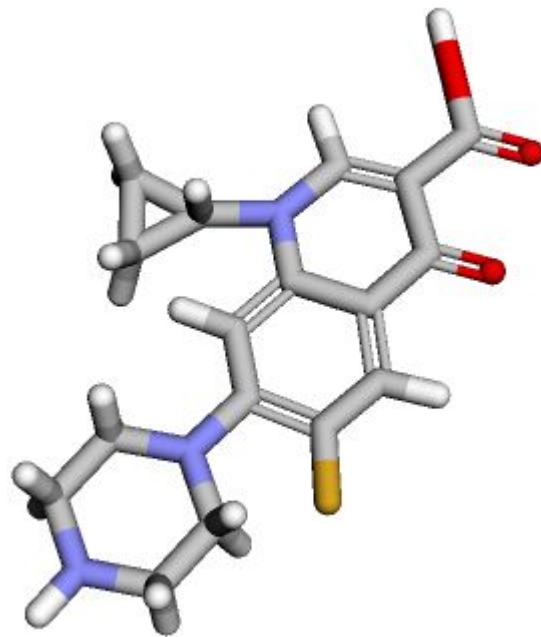


Johnson's Algorithm

- 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

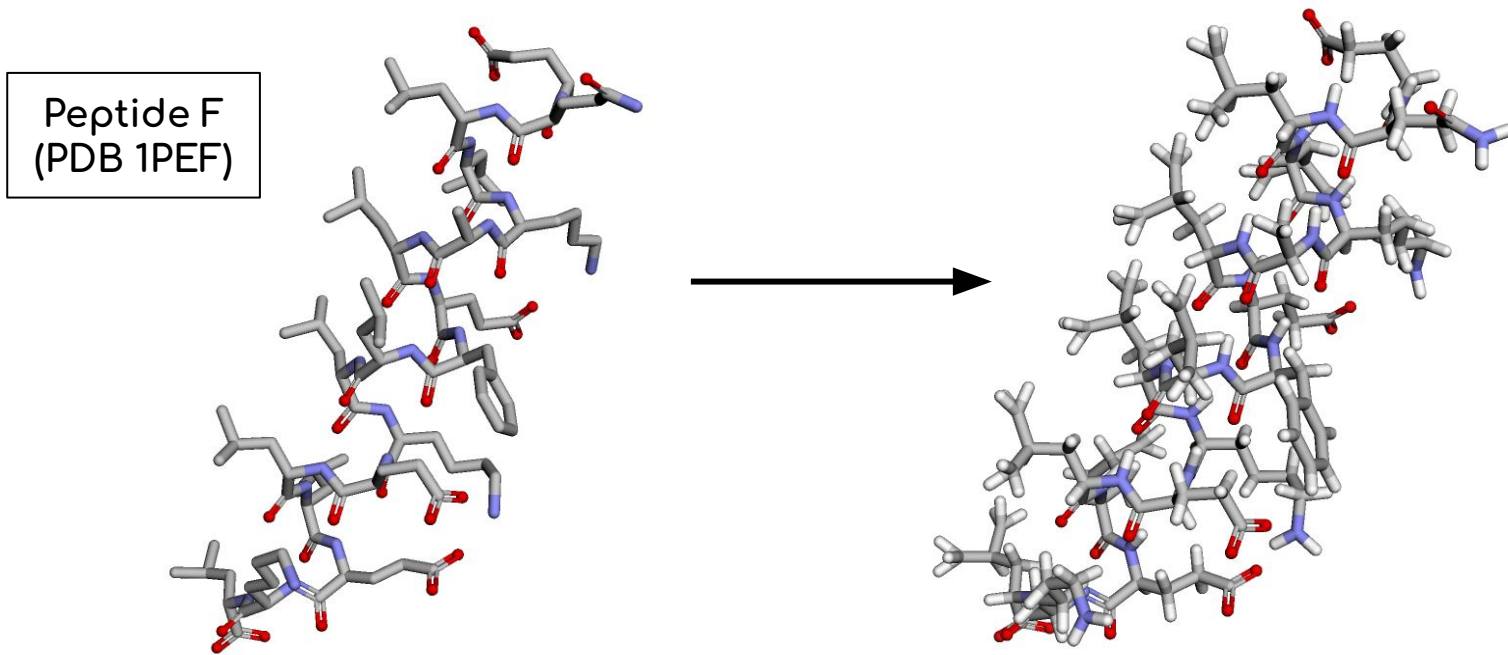


Ciprofloxacin structure made from:

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

Adding hydrogens

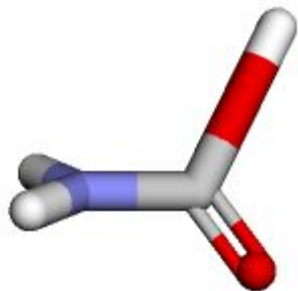
- PsiMol enables adding explicit hydrogen atoms to molecules



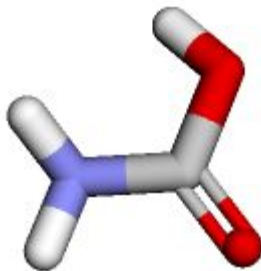
Integration with psi4

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

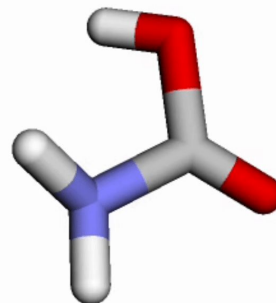
```
.from_smiles('NC(=O)O')
```



```
.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