

Automate Bulk Downloading of Molecular Properties and 3D Structures from
In a [previous blog post](#), I showed how to automate the downloading of 3D molecular structures and properties from [PubChem](#) using Python. Today's blog post is very similar but this time the database we'll be using is [ChemSpider](#) - a valuable resource for chemical information. In my experience, I've found that the structures on ChemSpider are more reasonable, especially if you're planning to perform quantum chemistry calculations.

Quick Recap: Why Automate?

If you've ever been faced with the task of gathering 3D structures and properties for a multitude of molecules, you know the manual process can be time-consuming and tedious. Python, with its versatile libraries, offers an efficient solution to automate this workflow and help you obtain the data for your machine-learning models.

Getting Started

Installing ChemSpiPy

Firstly, you need to ensure you have the necessary library installed. We need the [ChemSpiPy](#) python library. ChemSpiPy serves as a Python wrapper, enabling convenient access to the web APIs provided by ChemSpider. The objective is to furnish users with an interface for interacting with and querying the ChemSpider database through Python. This simplifies the development of programs capable of automating tasks that would typically be done manually on the ChemSpider website.

If you haven't already, install Chempipy using:

```
pip install chemspipy
```

Obtaining API Key

Unlike PubChem, which doesn't require any API or user account to obtain the data from their database, ChemSpiPy python library needs an API KEY to obtain any information from the ChemSpider database. To obtain one, [Register for a RSC Developers account](#) and then [Add a new key](#).

The web services provided by the Royal Society of Chemistry are presently accessible as part of an Open Developer Preview. Throughout this preview phase, users are granted the ability to make up to 1000 calls per month. If a higher allowance is needed, it is advised to reach out to api@rsc.org for further assistance.

The limit of 1000 called per month may be an issue for you if you need to make more calls for your purpose.

Python Script for downloading molecular structures and data

The following is a Python script for automating the download of molecular properties and 3D structures from ChemSpider.

```
from chemspipy import ChemSpider
import os
import pandas as pd

cs = ChemSpider('<API_KEY>')

# Define the list of molecules
molecules = [
    "methane",
    "propan-2-one",
    "2-acetyloxybenzoic acid",
```

```
"pentanal"
]

# Create a directory to save the SDF files
output_dir = "ChemSpider_molecule_structures"
os.makedirs(output_dir, exist_ok=True)

# Initialize lists to store information
molecule_names = []
common_names = []
formulas = []
molecular_weights = []
smiles_list = []

for molecule in molecules:
    print('Trying for ', molecule)
    compound = cs.search(molecule)
    if compound:
        print(molecule + ' found in ChemSpider database. Downloading...')
        compound = compound[0]
        print(compound.molecular_formula)
        print(compound.molecular_weight)
        print(compound.smiles)
        molecule_names.append(molecule)
        common_names.append(compound.common_name)
        formulas.append(compound.molecular_formula)
        molecular_weights.append(compound.molecular_weight)
        smiles_list.append(compound.smiles)
        # Write SDF to file
        molfile = open(os.path.join(output_dir, f'{molecule}.mol'), 'w')
        molfile.write(compound.mol_3d)
        molfile.close()

        print(f'Downloaded structure for {molecule}')
    else:
        print(f'No information found for {molecule}')

# Create DataFrame
data = {
    'Molecule Name': molecule_names,
    'Common Name': common_names,
    'Formula': formulas,
    'Molecular Weight': molecular_weights,
    'SMILES': smiles_list
}
df = pd.DataFrame(data)

# Export DataFrame as CSV and Excel files
df.to_csv('molecule_info.csv', index=False)
```

```
df.to_excel('molecule_info.xlsx', index=False)
print('All structures downloaded and information saved successfully!')
```

Using the Script

1. Insert Your ChemSpider API Key:
 - Before using the script, visit this link (<https://developer.rsc.org/user/register>) to obtain your API key.
 - Insert your API key in the script where it says `<API_KEY>`.
2. Define Molecules:
 - Edit the `molecules` list with the names of the compounds you're interested in.
3. Run the Script:
 - Execute the script, and watch as it automatically downloads 3D structures (in the directory `ChemSpider_molecule_structures`) and other properties (in files `molecule_info.csv`, and `molecule_info.xlsx`) from ChemSpider for the specified molecules.

Output

The output looks something like this

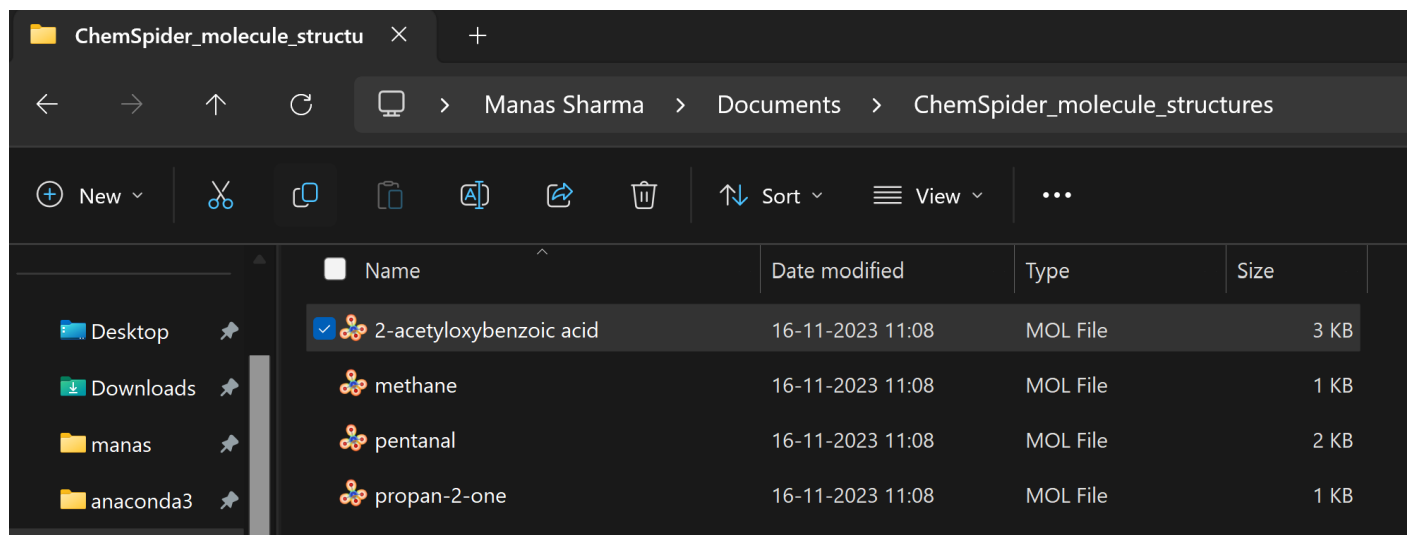
```
Trying for methane
methane found in ChemSpider database. Downloading...
CH_{4}
16.0425
C
Downloaded structure for methane
Trying for propan-2-one
propan-2-one found in ChemSpider database. Downloading...
C_{3}H_{6}O
58.0791
CC(=O)C
Downloaded structure for propan-2-one
Trying for 2-acetyloxybenzoic acid
2-acetyloxybenzoic acid found in ChemSpider database. Downloading...
C_{9}H_{8}O_{4}
180.1574
CC(OC1=C(C(=O)O)C=CC=C1)=O
CC(=O)OC1C=CC=CC=1C(O)=O
Downloaded structure for 2-acetyloxybenzoic acid
Trying for pentanal
pentanal found in ChemSpider database. Downloading...
C_{5}H_{10}O
86.1323
CCCCC=O
Downloaded structure for pentanal
All structures downloaded and information saved successfully!
```

The CSV file (`molecule_info.csv`) looks like this

Molecule Name	Common Name	Formula	Molecular Weight	SMILES
---------------	-------------	---------	------------------	--------

		Automate Bulk Downloading of Molecular Properties and 3D Structures from ChemSpider with Python and Chempipy 4		
methane	Methane	<chem>CH_{4}</chem>	16.0425	<chem>CC(=O)C</chem>
propan-2-one	Acetone	<chem>C_{3}H_{6}O</chem>	58.0791	<chem>CC(OC1=C(C(=O)O)C=CC=C1)=O</chem>
2-acetyloxybenzoic acid	Aspirin	<chem>C_{9}H_{8}O_{4}</chem>	180.1574	<chem>CC(=O)OC1C=CC=CC=1C(O)=O</chem>
pentanal	n-pentanal	<chem>C_{5}H_{10}O</chem>	86.1323	<chem>CCCC=O</chem>

and the molecular structures can be found in `ChemSpider_molecule_structures` folder



These .mol files can be visualized using any chemical file visualizer like [Jmol](#), [VESTA](#), [CrysX-3D Viewer](#), etc.

Potential Use Cases

This automated workflow can be a game-changer in various scenarios, including:

- Seeding molecular dynamics simulations with diverse starting structures.
- Benchmarking quantum chemistry methods on various organic molecule sets.
- Collecting data for machine learning projects.
- Populating an internal company database with molecules of interest.

Wrapping Up

By automating the bulk download of chemical information from ChemSpider, you save time in your research and projects. If you found my [previous guide on PubChem](#) useful, this ChemSpider script adds another tool to your arsenal.

Feel free to adapt and extend this script based on your specific needs. Happy coding, and let the molecular exploration continue!

If you have any doubts or suggestions then let me know in the comments section down below.

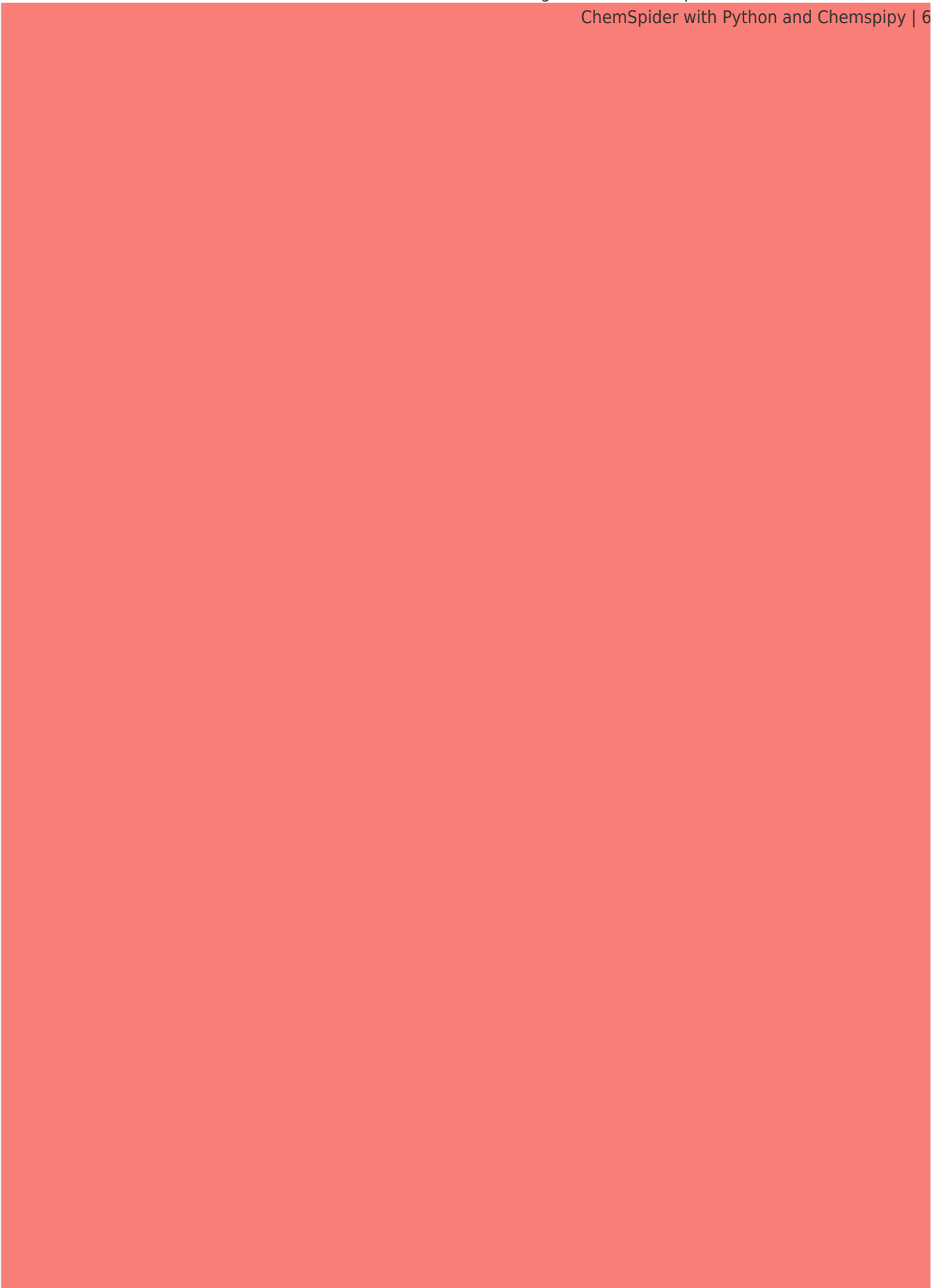


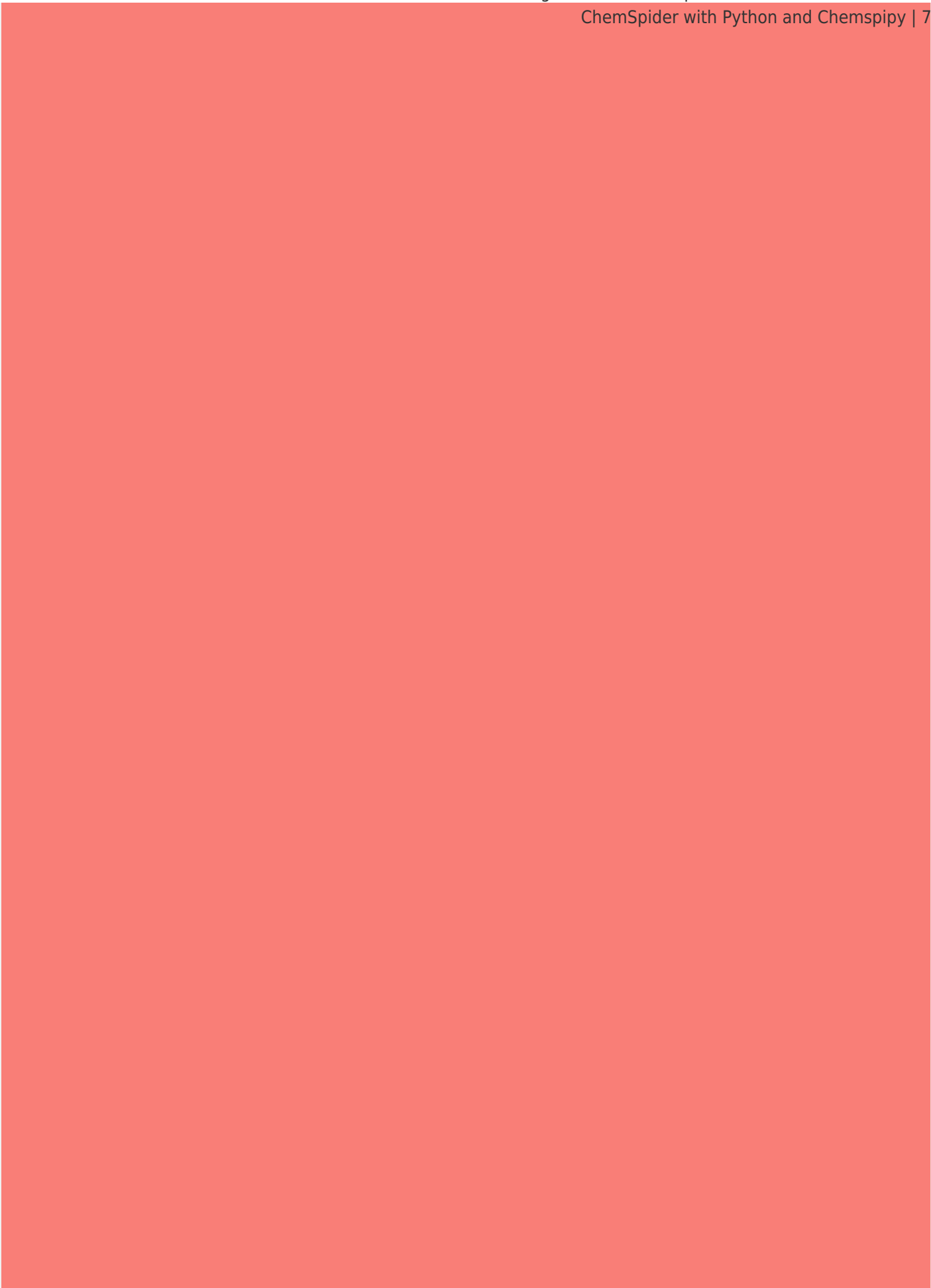
Manas Sharma

Automate Bulk Downloading of Molecular Properties and 3D Structures from

ChemSpider, PubChem, and CAS | 5
I'm a physicist specializing in computational material science with a PhD in Physics from Friedrich Schiller University Jena, Germany. I write efficient codes for simulating light-matter interactions at atomic scales. I like to develop Physics, DFT, and Machine Learning related apps and software from time to time. Can code in most of the popular languages. I like to share my knowledge in Physics and applications using this Blog and a YouTube channel.

manas.bragitoff.com/





Share this:

[Click to share on Facebook \(Opens in new window\)](#)

[Click to share on Twitter \(Opens in new window\)](#)

[Click to share on WhatsApp \(Opens in new window\)](#)

[Click to share on Pinterest \(Opens in new window\)](#)

[Click to share on Reddit \(Opens in new window\)](#)

[Click to share on LinkedIn \(Opens in new window\)](#)

[Click to email a link to a friend \(Opens in new window\)](#)

[Click to print \(Opens in new window\)](#)

[Click to share on Tumblr \(Opens in new window\)](#)

[Click to share on Pocket \(Opens in new window\)](#)

[Click to share on Telegram \(Opens in new window\)](#)

[wpedon id="7041" align="center"]