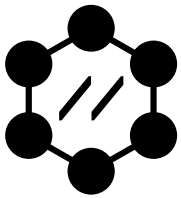


svatunek // lab

Computational Organic Chemistry

Why Chemistry Needs Python

Assistant Professor Dr. **Dennis Svatunek**

University of Malaya, 6th of November 2025

svatunek // lab

Computational Organic Chemistry

svatunek-lab.com

teaching.svatunek-lab.com

Research Talk: 11.11.2025 10 am

What is Computational Chemistry?

Computational chemistry is the science of **simulating chemical systems** on a computer.

It allows us to:

- **Model molecules and reactions** instead of (or before) doing experiments
- **Simulate** properties such as structure, energy, spectra, and reactivity
- **Explore** mechanisms, reaction pathways, and material properties
- **Design** new molecules, catalysts, and drugs using computation

What is Cheminformatics?

Cheminformatics combines chemistry, data science, and computer science to analyze, predict, and design chemical compounds using data.

It involves:

- **Representing molecules** digitally (e.g., SMILES, fingerprints, graphs)
- **Searching and comparing** chemical structures and properties
- **Predicting** reactivity, toxicity, or activity using **machine learning**
- **Managing and visualizing** large chemical datasets

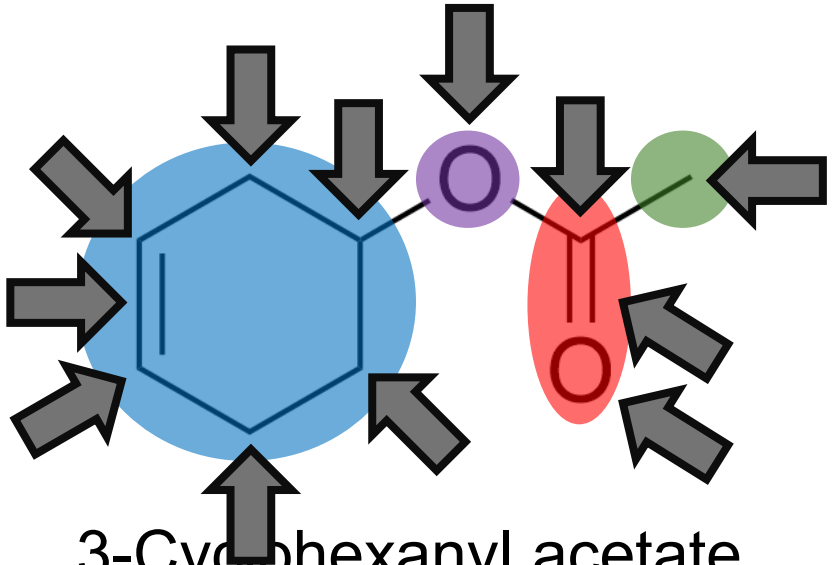
Why Python in Chemistry?

Python is widely used across scientific disciplines for its clarity, flexibility, and large ecosystem of tools.

It allows chemists to:

- **Automate** calculations, analyses, and simulations
- **Process and visualize** chemical and experimental data
- **Connect** quantum chemistry, molecular modeling, and AI tools
- **Build** custom workflows for research and teaching

Digital Molecules



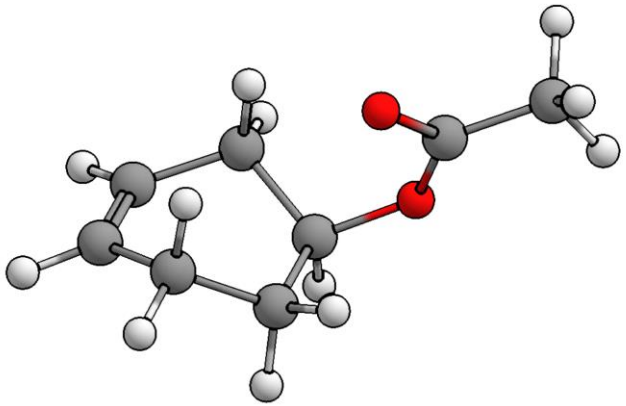
3-Cyclohexanyl acetate

Simplified Molecular Input Line Entry System

SMILES

Digital Molecules in 3D

XYZ format



```
22
Comment line. Can be blank
O       -1.446210000      1.270070000      0.747456000
C       -0.486601000      1.223970000      1.563790000
O        0.625008000      0.401471000      1.330890000
C        0.808628000     -0.429382000      0.178474000
C        1.005920000      0.426254000     -1.096010000
C        0.928128000     -0.439367000     -2.323250000
C       -0.050380100     -1.347130000     -2.424190000
C       -1.040170000     -1.492780000     -1.300800000
C       -0.314302000     -1.508160000      0.057636300
C       -0.516296000      2.085090000      2.780130000
H        1.764670000     -0.977125000      0.333702000
H        1.989370000      0.942079000     -1.060740000
H        0.216833000      1.202730000     -1.170510000
H        1.683420000     -0.360387000     -3.099050000
H       -0.093452100     -2.011200000     -3.281920000
H       -1.746640000     -0.635417000     -1.331250000
H       -1.625050000     -2.428280000     -1.430830000
H        0.140877000     -2.514170000      0.194326000
H       -1.064870000     -1.395000000      0.868704000
H       -0.203855000      3.115450000      2.511290000
H       -1.542680000      2.110940000      3.203100000
H        0.177496000      1.682650000      3.547680000
```

RDKit

RDKit is an **open-source cheminformatics toolkit** for working with chemical structures in Python. It provides the core functionality needed to represent, analyze, and visualize molecules.

It allows us to:

- **Read and write** molecules
- **Compute** properties and molecular descriptors
- **Generate** 2D depictions and 3D conformers
- **Search** and compare structures (substructure, similarity)
- **Prepare** data for **machine learning and modeling**

Practical Section 1

teaching.svatunek-lab.com/malaya1

*In this section, we explore how chemistry can be represented and analyzed directly in Python using **RDKit**.*

Machine Learning

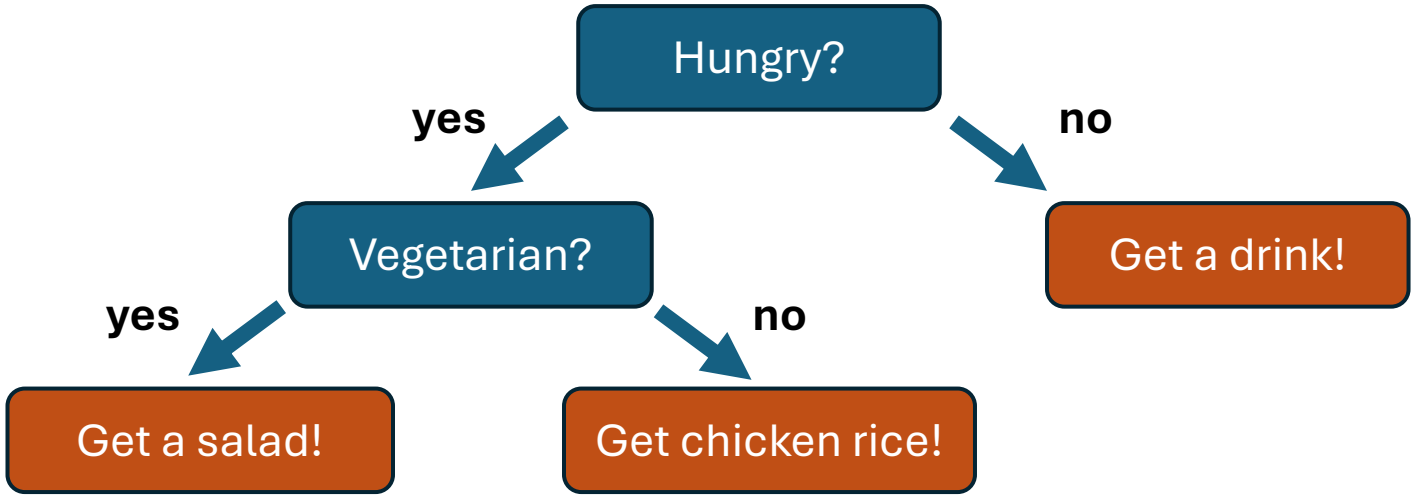
Machine learning connects **data** with **chemical properties and behavior**.By training models on known data, we can predict properties of new molecules.

It allows us to:

- **Correlate** molecular structure or descriptors with properties
- **Predict** activity or stability of new compounds before synthesis
- **Classify** molecules into functional or activity groups
- **Accelerate** discovery by learning from experimental or computed datasets

Decision Tree

A **decision tree** predicts an outcome by repeatedly splitting data based on feature thresholds, creating a series of if–then rules that lead to a final prediction at each leaf.



Decision Tree

A **decision tree** predicts an outcome by repeatedly splitting data based on feature thresholds, creating a series of if–then rules that lead to a final prediction at each leaf.

In regression it predicts the **average value** of samples in a leaf, while in classification it predicts the **most common class**.

Random Forest

A random forest combines the predictions of many individual decision trees, each trained on different subsets of data, to produce a more accurate and stable overall result.

Practical Section 2

teaching.svatunek-lab.com/malaya2

In this section, we learn how to make predictions from chemical data with basic machine learning tools.

Bonus: Quantum Chemistry in Python

PySCF brings *quantum mechanics* into Python, allowing us to compute chemical properties directly from first principles.
With only a few lines of code, we can run quantum-mechanical calculations, and analyze results.

It allows us to:

- **Compute** molecular energies, orbitals, and electron densities
- **Optimize** structures or scan reaction coordinates
- **Visualize** orbitals and charge distributions interactively

Bonus: Practical Section 3

teaching.svatunek-lab.com/malaya3

In this section, we explore how quantum mechanics can be used to calculate molecular structures, energies, and properties from first principles.

svatunek // lab

Computational Organic Chemistry

svatunek-lab.com

teaching.svatunek-lab.com

dennis.svatunek@tuwien.ac.at