

Environment setup sheet

Everything here is free. No licence and no paid database is required at any point.

What you need

- A computer running Windows, macOS or Linux with about 5 GB free and an internet connection for the data fetches.
- Python 3.9 or newer, and a notebook environment. Visual Studio Code with the Python and Jupyter extensions is free.
- Seven packages. pip install rdkit requests numpy pandas scipy scikit-learn matplotlib
- AutoDock Vina, installed separately, with Meeko or OpenBabel for file preparation.

Setting up the project

Step	Command or action
1 Create the project folder	Subfolders for data, output and notebooks
2 Create an isolated environment	python -m venv .venv then activate it
3 Install the packages	pip install rdkit requests numpy pandas scipy scikit-learn matplotlib
4 Verify the chemistry engine	from rdkit import Chem then Descriptors.MolWt of the molecule from CCO, which prints about 46
5 Freeze the versions when finished	pip freeze written to requirements.txt

Troubleshooting

Symptom	Usual cause	Fix
Import of RDKit fails	Wrong interpreter selected in the notebook	Select the kernel from the project environment
Many compounds fail to fetch	Service timeouts rather than bad names	Print the status code, add retries and a cache
Numbers change between runs	Package version changed	Recreate the environment from requirements.txt
Docking program not found	Vina not on the path	Call it by its full path or add it to the path

Based on the ERAS Africa Molecular Docking using Python course, 2026, and the study "Target Protein Pocket Volume and Antiretroviral Drug Physicochemical Properties" (Hailekiros, 2026) with its reproduction guide. Tool documentation from RDKit, AutoDock Vina, Meeko, RCSB PDB, PubChem and ProteinsPlus. Free to copy and adapt with attribution.