# mol2crystal: Automatic Crystal Structure Generation Tool from Molecular Input

## Overview

mol2crystal is a Python-based software tool that takes a single molecular structure (in MOL format, specifically named precursor.mol) as input and automatically generates and evaluates candidate molecular crystal structures using space group symmetry. Key features include:  
- Exhaustive crystal structure generation using space groups (1–230)  
- Automatic detection of atomic overlaps  
- Support for high-precision post-processing tools such as DFTB+, xtb, and DFT-D3  
- Fully automated structure selection and density evaluation

## Environment Requirements

Python version: 3.8 or higher  
Recommended packages (verified versions):  
pip install ase==3.22.1 scipy==1.13.0 psutil==7.0.0  
pip install pymsym==0.3.4 spglib==2.6.0

## Directory Structure (After Execution)

molecular\_files/: Place input molecular file (e.g., precursor.mol)  
valid\_structures/: Accepted structures (VASP format)  
optimized\_structures\_vasp/: Post-processed optimized structures  
structure\_vs\_energy.txt: Records of density, volume, atom count, etc. for each structure

## Usage Instructions

1. Place the molecular structure in molecular\_files/precursor.mol  
2. Run the following command:  
 python3 mol2crystal.py  
3. Results will be output to valid\_structures/, and density information will be recorded in structure\_vs\_energy.txt

## User Settings (Configurable at the Beginning of the Script)

user\_margin: Margin around the molecule (Å), e.g., 1.70  
user\_nmesh: Number of divisions for rotation angles, e.g., 3  
user\_overlap\_scale: Threshold scale for atomic distance, e.g., 0.80  
user\_max\_depth: Depth of space group expansion (based on physical similarity), e.g., 1  
user\_skipping\_n\_molecules: Exclude structures with too many molecules, e.g., 100

## Post-processing (Examples)

High-precision optimization can be performed on selected structures using the following scripts:  
python3 postprocess\_dftb.py  
python3 postprocess\_xtb.py  
python3 postprocess\_siesta.py

## Output Example (structure\_vs\_energy.txt)

valid\_structures/POSCAR\_theta\_0\_phi\_0\_sg\_62 0.000000 1.234 48 1234.56

## Background and Significance

mol2crystal aims to automate and ensure reproducibility in molecular crystal structure prediction. It contributes to education, research, and industrial applications as a fully open and reproducible structure prediction tool.

## Code Structure and Key Scripts

mol2crystal.py: Main script for automatic crystal structure generation from molecular input  
select\_data.py: Structure selection based on energy, density, etc.  
postprocess\_\*.py: High-precision optimization using DFTB+, xtb, SIESTA  
run.sh: Bash script to manage overall execution with logging

## Processing Flow

1. Read and center the molecule  
2. Analyze point group and extract candidate space groups  
3. Rotate molecule and define unit cell  
4. Apply space group and generate structures  
5. Check for atomic overlaps  
6. Save structure and evaluate density

## Key Functions

rotate\_molecule(): Rotate molecule by θ and φ angles  
has\_overlap\_neighborlist(): Quickly detect atomic overlaps  
adjust\_cellpar\_by\_spacegroup(): Adjust unit cell parameters according to space group  
get\_primitive\_cell(): Convert to primitive cell using spglib  
density\_calc(): Calculate density from mass and volume and log the result  
get\_all\_subgroups(): Recursively obtain strict subgroups of point groups  
expand\_physical\_supergroups(): Expand physically similar supergroups to a specified depth

## Design Philosophy

- Strict yet flexible handling of point group and space group correspondence  
- Emphasis on automation and reproducibility in structure generation  
- Designed with integration to post-processing workflows in mind