**Scaffold Decorator**

Scaffold decorator consists of two python scripts: one for the collection of R-groups and other one for generation of new molecules.

**Requirements**

* Python 3
* RDKit version 2022.03.5 (or higher)
* Regex
* Pandas
* Numpy

**Folder Structure**

**scaffold\_decorator/**

**├── collect\_r\_groups/**

**│ ├── collect\_r\_groups.py**

**│ ├── example\_molecules.smi**

**│ └── r\_groups\_output.smi**

**│**

**└── generate\_new\_molecules/**

**│ ├── make\_new\_molecules.py**

**│ ├── r\_groups\_unique.smi**

**│ └── example\_scaffolds.smi**

**│**

**└── other\_files/**

**├── bioinfo\_db.smi**

**├── chembl\_clinical.smi**

**Step 1: R-Group Collection**

Run the script **‘collect\_r\_groups.py’**, providing molecules and scaffolds as an input to generate the ‘**r\_groups.smi’** file

**python collect\_r\_groups.py -m example\_molecules.smi -c example\_core\_scaffolds.smi**

**Input Files:**

**example\_molecules.smi**: Contains existing molecules in SMILES format.

**example\_core\_scaffolds.smi**: Contains core scaffolds in SMILES format.

**Output**:

* **r\_groups.smi**: Contains the R-groups identified from the decomposition

### Step 2: Generation of New Molecules

Run the script ‘**make\_new\_molecules.py’**, providing the scaffolds with the substitution points, a collection of R groups, and the desired number of R group combinations to generate a file **‘products.smi’** that contains novel molecules

**python make\_new\_molecules.py -r r\_groups\_unique.smi -c example\_scaffolds.smi -n 1000**

1. **Input Files**:
   * **r\_groups.smi**: Contains the R-groups for decoration.
   * **example\_scaffolds.smi**: Contains scaffolds to be decorated.
   * Make sure that scaffolds 2-4 substitution points and > 4 substitution points are separated
   * Number of substitution points and number of R-group combinations should be provided.

**Output**:

1. **new\_molecules.smi**: Contains the newly generated molecules

**Note**

* Use R-groups with multiplicities to ensure the good distribution of different R-group types
* Make sure to filter the molecules after generating them. The filtering rules are provided in Supplementary Materials 5.

### Summary

* Ensure that your environment has Python 3, RDKit 2022.03.5 (or higher), Pandas, Regex, and NumPy.
* Follow the directory structure and place input files accordingly.
* Run the provided scripts to collect R-groups and generate new molecules.