EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

BIO-4372
CHEMINFORMATICS
Winter Semester 2022-23

Dr. Philipp Thiel
Institute for Bioinformatics
and Medical Informatics

Assignment Sheet 5

Submission deadline: November 28, 2022, 9 a.m.

A5.1: Murcko Scaffolds: Part I [5 points]

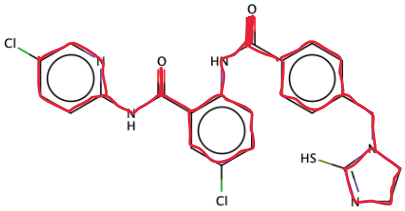
The concept of molecular scaffolds (also called frameworks) is very important in medicinal chemistry and drug design. Various scaffold definitions exist and all aim to describe the core structure of a molecule. One very prominent example is the scaffold definition given by Bemis and Murcko [1,2]. This type of scaffold is usually referred to as the `Murcko Scaffold` or `Murcko Framework` of a given molecule. Please familiarize yourself with the concept of Murcko Scaffolds and explain it briefly in your report (maximally 200 words, images are allowed).

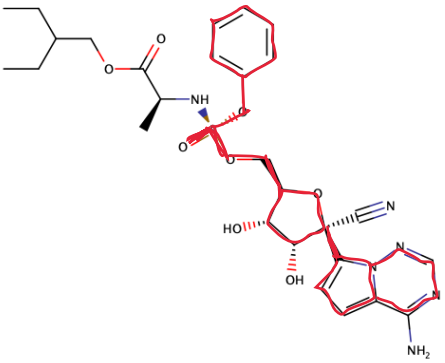
[1] `https://doi.org/10.1021/jm9602928`

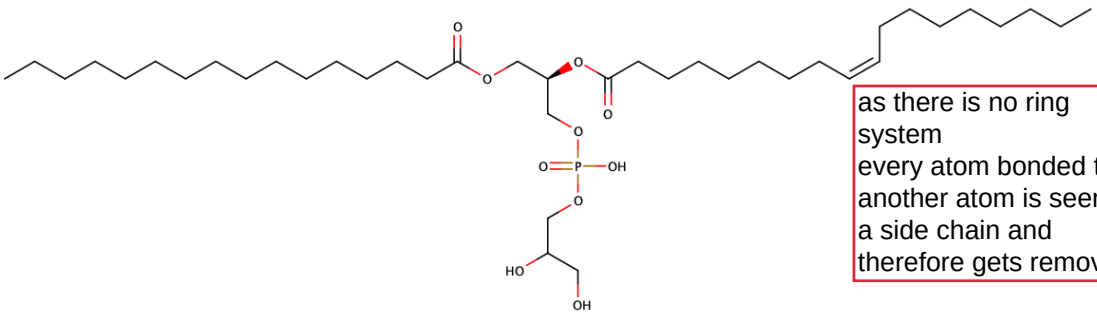
[2] `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3180201`

A5.2: Murcko Scaffolds: Part II [5 points]

Please highlight the atoms and bonds that belong to the Murcko Scaffolds of the following molecules:







as there is no ring system
every atom bonded to another atom is seen as a side chain and therefore gets removed

A5.3: RDKit and Python in KNIME [10 points]

Cheminformatics is a discipline that requires skills in various programming languages, software libraries, and software tools because the functionalities that are required to solve complex tasks are usually distributed over a large number of sources. The ability to work in this heterogeneous environment and to integrate these bits and pieces into useful tools is - in my humble opinion - a key competence in this field. One real-world example is the ability to execute RDKit Python scripts within KNIME, which you can practice now. The task is to read the molecules from `chin-a5-compounds.csv`, convert them into RDKit molecules and to implement a script that generates the Murcko Scaffolds of these molecules, which are added as a new column to the output table. Please follow the described steps:

1. Make sure that the `KNIME Python Integration` is available
2. Read the SMILES from the given file and convert them to RDKit molecules
3. Find a suitable node to implement and execute Python code
4. Make sure that you have configured a Conda environment with installed RDKit
   (Tipps: KNIME preferences; you can refer to your already existing environment!)
5. Find out how to add a new column to the output table of the scripting node.
   (Tipp: input and output tables are Pandas DataFrames!)
6. Implement Murcko Scaffold generation in this node and add them as column 'MurckoScaffold'
   - The algorithm is rather straightforward, really!
   - Have a close look on the degree of 'side chain atoms'
   - And you should familiarize with removal of atoms in RDKit
   - Run into a range error? The big web-oracles will help you, sure!

Submission

1. Report with solutions for A5.1 and A5.2
2. Your properly exported KNIME workflow for A5.3

Please bundle the files in a tar-archive to simplify uploading to ILIAS.

▷ Please use Slack to discuss problems in the first place
▷ If you have confidential questions, don't hesitate to drop by or write an e-mail