EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

BIO-4372
CHEMINFORMATICS
Winter Semester 2022-23

Dr. Philipp Thiel
Institute for Bioinformatics
and Medical Informatics

Assignment Sheet 2

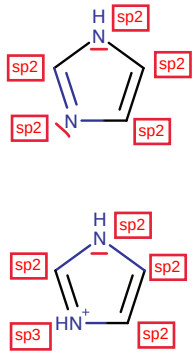
Submission deadline: November 7, 2022, 9 a.m.

Please use the report template for this assignment sheet.

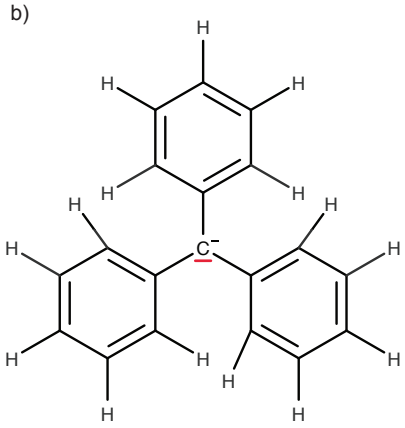
A2.1: Practicing Some Chemistry [6 points]

For the molecules shown in the figure below you shall determine important chemical properties. For the molecules shown in a) please add missing lone pairs and determine the hybridization of all atoms. For the molecule shown in b) please add missing lone pairs. For molecules shown in c) please add missing lone pairs and determine the hybridization of the oxygen atoms. Add a figure to your report where the required annotations have been made. Feel free to use some software for annotation purpose or a photo of your manual annotations.

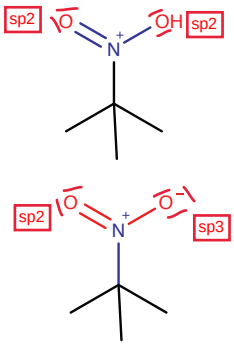
a)



b)



c)



A2.2: HTS Data Analysis [14 points]

High-throughput screening (HTS) is the most important experimental procedure in the lead ID phase to search for primary hits. Standard HTS screening plates consist of 384 wells with 16 rows and 24 columns. In every well, the same assay is performed for different compounds (samples) to identify candidates that show a significant effect. If available, a positive control (PC), which is a reference compound with known and strong effect, and a negative control (NC), which is the assay without added compound, are also tested on every plate. The raw data of an HTS campaign is usually one file per screening plate that contains the measured raw values. These data has to be analyzed in order to pick primary hits.

The provided archive `material.tar.gz` contains a folder with 500 HTS raw-data files of 384-well screening plates. Column 1 always contains the PC and column 24 the NC. Plate columns 2-23 contain the samples, i.e. the tested compounds of your screening library. Each raw data file has the following columns: (1) project name, (2) plate ID, (3) row ID, (4) column ID, (5) sample ID, (6) raw measurement values. PC and NC have the sample

ID `ctrl`. The provided file `screening_library.csv` contains the SMILES of the screening library compounds.

Your task will be to process and analyze the given HTS data in order to select the primary hits. Tipp: browse through your KNIME nodes to get an idea about the wealth of helpful extensions.

1. Create a new workflow with the name `chin-a2-<lastname>-<firstname>`.
2. Read all HTS files into a single table.
3. Visualize screening plates using heat maps.
   Tipp: have a look at the KNIME extension '`HCS Tools`'!
4. Describe the Z-Prime factor for HTS quality control and evaluate your screening plates using this parameter.
5. Please identify plates with low quality, report their total number, and justify your reasoning.
   Tipp: looking at heat maps can be helpful!
6. Please split the data set into low-quality and high-quality plates.
7. To select hits from the high-quality plates, please use *normalized percent inhibition* (NPI) normalization.
   As a hit picking criterion use NPI $\geq 70.0\%$.
8. Hit selection from the low-quality plate requires an appropriate normalization method.
   The latter should be chosen based on the identified reason for insufficient quality of these plates
   and from the methods discussed in the lecture. To select hits please use a meaningful hit picking criterion.
9. Merge both hit lists.
10. Read the screening library file and join the compound data onto your hit list.
11. Sort the final hit list by sample ID and write the results into a tab-separated-values (TSV) file with the following columns: (1) sample ID, (2) molecular weight, (3) #h-bond donors, (4) #h-bond acceptors, (5) SlogP. No. 2-5 are so-called descriptors that can be calculated from the molecules.

Describe your HTS data analysis briefly but precisely.
Include only exemplary figures to support your results and findings.

Submission

1. Your report in PDF format.
2. Your properly exported KNIME workflow.

▷ 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