

Exercise Sheet 10

Exercise 36 Molecular Description Languages

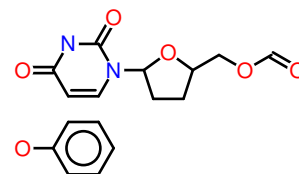
a) Draw the molecule that is described by the following SMILES string:

S2c1c4c(ccc1N(c3c2cccc3)C(=O)C)cccc4

b) Draw the molecule that is described by the following SLN string:

NH2C[1]:N:C(:C(:N:C:@1C(=O)OCH3)C(=O)C[2]:CH:CH:CH:CH:CH:@2)NH2

c) Construct a SMILES description of this molecule:



d) Find at least three different SMILES descriptions of phenol, that is, of this molecule:

Exercise 37 Molecular Description Languages

Draw the molecule that is described by the following SDfile:

```
4728
      1016 111283D
      4728
9  9      0      1 V2000
-0.0965  1.3884  0.0104 N  0  0  0  0  0  0  0  0  0  0  0  0
 1.8297 -0.2821 -0.0166 S  0  0  0  0  0  0  0  0  0  0  0
 1.0302  2.1371  0.0038 N  0  0  0  0  0  0  0  0  0  0  0
 0.1607  0.0664  0.0009 C  0  0  0  0  0  0  0  0  0  0  0
 2.1519  1.3924 -0.0108 C  0  0  0  0  0  0  0  0  0  0  0
 3.3971  1.9238 -0.0196 N  0  0  0  0  0  0  0  0  0  0  0
-0.5873 -0.6698  0.0040 H  0  0  0  0  0  0  0  0  0  0  0
 3.5072  2.8871 -0.0156 H  0  0  0  0  0  0  0  0  0  0  0
 4.1746  1.3453 -0.0300 H  0  0  0  0  0  0  0  0  0  0  0
1  4  2  0  0  0  0
1  3  1  0  0  0  0
2  4  1  0  0  0  0
2  5  1  0  0  0  0
3  5  2  0  0  0  0
4  7  1  0  0  0  0
5  6  1  0  0  0  0
6  8  1  0  0  0  0
6  9  1  0  0  0  0
M  END
$$$$
```

Exercise 38 Subgraph Isomorphisms

- How often does the fragment C-C-C occur in phenol (see Exercise 36d)?
In other words: How many subgraph isomorphisms exist?
- Does phenol (see Exercise 36d) possess an automorphism that is not the identity?
How many different automorphisms does phenol possess?
- Construct the set of all connected subgraphs of phenol (see Exercise 36d)!
- Find all bonds in phenol (see Exercise 36d) that are bridges!
Are any of these bridges proper bridges?
- Find all bonds in the graph/molecule from Exercise 36c) that are bridges!
Which of these bridges are proper bridges?

Exercise 39 Searching for Frequent Subgraphs

- a) Find all frequent (sub)graphs containing sulphur for a minimum support $s_{\min} = 3$ in the graph database (SMILES format) shown on the right:
- | | |
|-------------------------|-----------------------|
| <chem>CCS(O)(O)N</chem> | <chem>CCS(=N)N</chem> |
| <chem>CCS(O)(C)N</chem> | <chem>CS(=N)N</chem> |
| <chem>CS(O)(C)N</chem> | <chem>CS(=N)O</chem> |
- b) Why is it more difficult to avoid redundant search when searching for frequent (sub)graphs than for frequent item sets? What are the main problems?
- c) What is the purpose of constructing code words of (sub)graphs? What is a canonical code word? What do we actually need?
- d) How do we assign a unique parent (sub)graph based on a canonical code words for a (sub)graph? What information from the canonical code word do we use?