# Seminar 1 - Basic operations with small molecules (ligands)

Author(s): Jan Zitko ([jan.zitko@faf.cuni.cz](mailto:jan.zitko@faf.cuni.cz)) and Martin Juhás ([juhasm@faf.cuni.cz](mailto:juhasm@faf.cuni.cz))

Updated 6th March 2023

## REQUIRED SOFTWARE

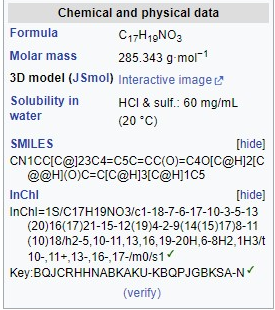
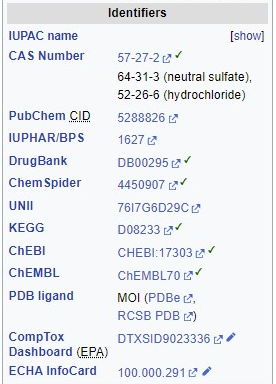
* Chimera 1.16:  
  (comprehensive manual: [Introduction to UCSF Chimera – Part 1 – (marpat.github.io)](https://marpat.github.io/chimera-intro.html))
  + download from <https://www.cgl.ucsf.edu/chimera/download.html>

## THEORY AND INTRODUCTION

### How to obtain small molecules:

Wikipedia: (e.g. <https://en.wikipedia.org/wiki/Morphine>) contains identifiers and links to several databases, e.g.:

* CAS – [Chemical Abstracts Service Registry](https://www.cas.org/cas-data/cas-registry) (202 million organic and inorganic substances, including alloys, coordination compounds, minerals, mixtures, polymers, and salts disclosed in publications since the early 1800s)
* <https://pubchem.ncbi.nlm.nih.gov/> (all-purpose chemical database)
* <https://go.drugbank.com/> (available drugs with properties)
* <https://www.ebi.ac.uk/chembl/> (compounds and their biological properties)



Specifically for computational (docking) purposes, you can use the ZINC database as the source of your ligands. ZINC15 contains approx. 389M purchasable compounds (for-sale subset, 2022, <https://zinc15.docking.org/>).

The most common chemical identifiers are SMILES ([Simplified molecular-input line-entry system – Wikipedia](https://en.wikipedia.org/wiki/Simplified_molecular-input_line-entry_system)), InChI ([International Chemical Identifier – Wikipedia](https://en.wikipedia.org/wiki/International_Chemical_Identifier)). SMILES and InChI are structural identifiers; that is, they encode the compound’s structure (atom types, connectivity, stereochemistry) but do not contain 3D coordinates.

The most commonly used **file formats** to store chemical structures (2D or 3D) are XYZ, SDF/MOL, MOL2, PDB…([Chemical file format – Wikipedia](https://en.wikipedia.org/wiki/Chemical_file_format)). All these formats are text files and human-readable (check for yourself).

### Translators or convertors of chemical formats

Online SMILES Translator and Structure File Generators: <https://cactus.nci.nih.gov/translate/>

OpenBabel online: <http://www.cheminfo.org/Chemistry/Cheminformatics/FormatConverter/index.html>

Most common offline tool: Open Babel ([Open Babel](http://openbabel.org/wiki/Main_Page))

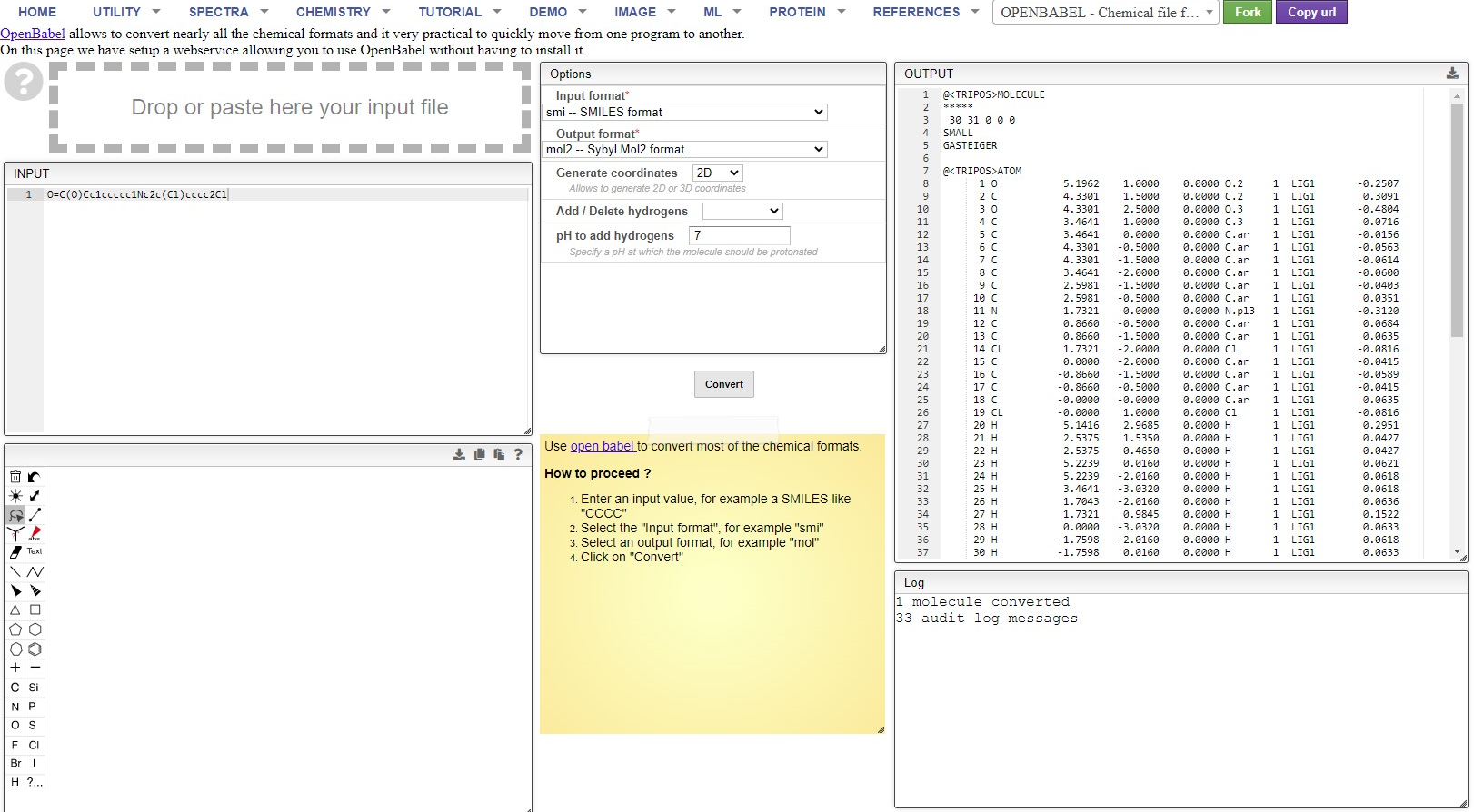
### Software for chemical drawing

Draw/build yourself within chemical software: **ChemDraw** (PerkinElmers Informatics), **ChemSketch** (ACD Labs), **MarvinSketch** (ChemAxon), **Avogadro** (Avogadro Chemistry), **Chimera** (UCSF)…

## PRACTICAL TRAINING

### Task 1 – Manipulation of small molecules in Chimera

1. Get the structure of **diclofenac**. Use SMILES found on Wikipedia and translate them, e.g. using OpenBabel translator to **2D structure** in **mol2 format** and save to file:



1. Open mol2 file in a text editor (e.g. notepad.exe) and observe the structure of the mol2 format. Indicate below, what structural information is present. (*HINT: look at the @ flags*)

**Answer:**

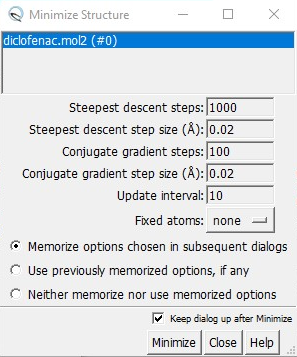
NOTE: When saving the mol2 file in the text editor, make sure that you use the encoding ‘UTF-8’ and not ‘UTF-8 with BOM’ or any other. With incorrect encoding, the opening of the mol2 file by Chimera in the next step will fail.

1. Then start Chimera and open the mol2 file (**File-Open...**). Observe the structure in Chimera. (Tutorial for Chimera: <https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/frametut.html> )

**Basics**: rotate - Left mouse button (MB); zoom - Right MB / wheel; translate - Middle MB

1. Prepare the ligand:

* Correct protonation state (remove carboxylic hydrogen)
* Add partial charges (**Tools-Structure Editing-Add Charge**)
* Show partial computed charges as labels (**Actions-Label-Other-Label with attribute-Charge**) to verify that charges were computed
* Hide the labels (**Actions-Label-Off**)
* Minimize the structure (**Structure Editing-Minimize Structure**). Use the following settings:



When asked about the charges for non-standard residues, select **AM1-BCC** (see e.g. [Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation - PubMed (nih.gov)](https://pubmed.ncbi.nlm.nih.gov/12395429/))

(Try running the minimization multiple times until no changes in the structure’s conformation are observed.)

1. Measure (**Tools-Structure Analysis-Distances**, select atoms by Ctrl+Shift+LMB):  
   (*Note:* Select atoms first, then calculate. Define plane by selecting 3 atoms and click **Define plane…** To measure an angle between two planes, select both planes in the table and check the value below the table). Report the measured values in the following table – do not forget to enter the correct units.

**Fill in the measurements:**

|  |  |
| --- | --- |
| The shortest distance between amine hydrogen and carboxylic oxygen: |  |
| The length of a C-C bond in the benzene ring |  |
| The angle O-C-O of the carboxylic group: |  |
| The C-C-C angle in the benzene ring: |  |
| The angle formed by the two benzene rings (defined as two planes): |  |

1. Experiments with different visualizations (e.g. **Actions-Atoms-Bonds-Stick or Sphere…**)
2. Save the final 3D structure as PNG image.
3. Save the final 3D structure as mol2.

### Task 2 – Prepare the molecule of diazepam

1. Get the SMILES string and use it as the input for a build-in conversion in Chimera (**Structure Editing-Build Structure**)

Note: Even though the structure generated by this procedure is in 3D, you still need to compute charges and minimize the structure to obtain the lowest energy conformation.

1. Save the final 3D structure as PNG image.
2. Save the final 3D structure as mol2.

## TAKE-HOME-MESSAGE

* Working with small molecules (ligands) for computational design requires:
  + Obtaining the 2D structure
  + Calculating 3D coordinates (in some programmes done automatically)
  + Assigning correct protonation state
  + Calculating and assigning partial charges
  + Energy minimization
  + Saving the results into a suitable chemical format file

## SUBMISSION CHECKLIST

As a result of this practical training, you are supposed to submit:

* This document with the **answered questions** and the table containing the measurements
* Two PNG files (diclofenac and diazepam) and three mol2 files (diclofenac 2D and 3D, diazepam 3D). For your convenience, the instructions to save (and submit) a file are always underlined in this document.