Introduction to Computational Chemistry
Exercises Part 3

Eno Paenurk & Patrick Finkelstein & Felix Pultar

May 10, 2023

Goals:

- Visualize xyz files

- Optimize structures with `xtb`

- Convert structures from ChemDraw to xyz (`smi2xyz.sh`)

- Resolve some errors that can occur

Note: we will only deal with a single 3D structure per compound. For your research, don't forget to consider conformers where necessary. There is a tool for systematic conformer searching called `crest` (link given in the Handout). We will not cover it here, but its general usage is similar to that of `xtb`.

---

1 Installing `smi2xyz.sh`

We will first guide you through setting everything up for today's calculations. You should already have `xtb` installed (cf. Week 2). You will additionally need to install Open Babel to your `icc-2023 conda` environment. If you're using Euler, this is already set up for you. If you're using Grace, load the `miniconda` module and run the following commands:

```
conda activate icc-2023
conda install -c conda-forge openbabel
```

To set up SMILES-to-xyz (`smi2xyz.sh`), update your ICC materials from the `git` repository as shown last week (cf. Section 2 in Ex. 2). Copy the `smi2xyz.sh` code from the code subdirectory to your `bin` directory (~/bin), and give it executing permissions. Now you can call `smi2xyz.sh` in any directory. We will learn its usage in Section 5.

2 Visualizing structures

Before we start creating structures, let us briefly check some of the options to visualize the Cartesian coordinate files. Start by creating a file called h2o.xyz in a subdirectory `h2o` in your Part 3 directory:

```
3
This second line is used for comments. It can be left empty, but it must exist.
O  0.0  0.0  0.0
H  1.0  0.0  0.0
H  0.0  1.0  0.0
```

The format of the xyz file is explained in the Handout.

There are two options to visualize this structure: on your computer with Avogadro/Chemcraft or in the terminal with Molden.

> **Avogadro/Chemcraft**
>
> Use your file browser (MobaXterm or FileZilla) to navigate to the directory your xyz file is at, right-click on the file and choose "Open with…" (MobaXterm) or "View/Edit" and then "Use custom program" (FileZilla). In both cases, you will then have to locate either the Chemcraft or Avogadro executable on your computer. If you make either of the programs the default one for opening xyz files, the xyz file will be loaded automatically with that program if you double-left-click on it (Chemcraft), or right click and "View/Edit" (FileZilla).
>
> In Chemcraft, you can change the visualization options under the "Display" menu. You can also customize your own visualization style (the CPK styles are a good starting point for that). To save the selected style as default, do `Display-Customize-Save all-Close`.
>
> In Avogadro, you can change the visualization settings by opening the "Display Settings…" (Avogadro 1) or "Display types" (Avogadro 2) tab. The changes are saved automatically.

> **Molden (Euler only)**
>
> Start by copying the `molden` executable from the ICC directory to your `bin` directory:
>
> ```
> cp /cluster/project/workshops/icc/icc-2023/software/molden6.9/bin/molden ~/bin/
> ```
>
> You can then call `molden` on an xyz file, which will open the Molden window:
>
> ```
> molden h2o.xyz
> ```
>
> As you can see, it is not the prettiest visualizer out there, but it can be handy for quickly checking a structure, or for working with files of specific (old) format. You can use the blue Molden control menu to turn off the shade and change the visualization mode to Solid-Sticks for the usual balls-and-sticks model. To exit, either click on the skull icon or press `Ctrl-C` in the terminal.
>
> For Mac and Windows users: if you use any other terminal than MobaXterm, you need to install an X11 forwarding application to visualize Graphical User Interfaces (GUIs), such as Molden. For example, XQuartz for Macs or Xming for Windows. Additionally, you need to use the `-Y` flag with the `ssh` command when accessing Euler:
>
> ```
> ssh -Y username@euler.ethz.ch
> ```

3 Interactive jobs

Converting or optimizing just a few structures to xyz files is easiest done interactively, i.e., not by submitting the calculations to the job queue. For new types of structures, this also enables you to first make sure that the SMILES conversion is working as wanted, or that the optimization leads to a reasonable result. As we said, **you should not run resource-intensive[1] commands on the login node**. Alternatively, you can apply for an interactive bash terminal on a compute node.

3.1 Euler

Apply for an interactive job with:

```
srun --pty --x11 bash
```

You will then be put in the queue for the requested resources. As mentioned last time, the default resources are 1 core and 4 hours. You can ask for 4 cores, for instance, by:

```
srun --pty --x11 -n 4 bash
```

---

[1]What is considered resource-intensive is up for debate, but we generally draw the line roughly at: more than 1 core; more than 4 GB expected RAM usage; longer than 1 minute.

You will see the message:

```
srun: job XXXXX queued and waiting for resources
```

with XXXXX being the job ID, followed by (once the job is running):

```
srun: job XXXXX has been allocated resources
```

3.2 Grace

The overall workflow is similar to Euler. The command for requesting an interactive job is:

```
salloc --x11
```

As on Euler, this command can be further decorated to give more resources. For example, you can ask for 4 cores and 4 GB of RAM for 4 hours by:

```
salloc --x11 -n 4 --mem=4G -t 4:00:00
```

3.3 General

> The `--x11` flag enables graphics forwarding, so you can use graphical software (like Molden or plotting tools). If you don't wish to use any graphical software during the interactive job, you can skip the `--x11` flag.

You have to wait for the resources to become available, which usually takes a minute or few (depending on your priority and the current work-load on the cluster). Once your interactive job is accepted, you will be able to use the terminal to run commands on the compute node for the specified amount of time.

Once you are done with the calculations, you should stop the interactive job to free up the resources. To stop the interactive job, type:

```
exit
```

> If you want to check whether you're on a login node, you can type:
>
> ```
> hostname
> ```
>
> If the returned string includes the word "login" (Euler) or "graceX.grace" (Grace), you're on a login node. You can also check that you have an ongoing bash job by typing:
>
> ```
> squeue
> ```

4 Optimizing geometries

You can optimize any xyz file with `xtb` to find the closest minimum energy structure on the potential energy surface. You can check the available `xtb` keywords by typing:

```
xtb -h
```

For optimizing the structure starting from an xyz file called filename.xyz, the usage is as follows:

```
xtb filename.xyz --opt
```

This will run the default optimization procedure. Some more options are explained in the Handout.

Once you have run the calculation, several files are created in the directory:

File	Description
charges	atomic partial charges
wbo	Wiberg bond orders
xtbopt.log	trajectory of the optimization
xtbopt.xyz	the optimized structure
xtbrestart	file for restarting calculations
xtbtopo.mol	xtb-specific topology file

Important: as the files are always given the same name, they will be overwritten if you run several calculations in the same directory. In addition, `xtb` will try to read the `xtbrestart` file, which may interfere with your new calculations. You should either change the filenames in between the calculations or have a separate directory for each calculation.

For now, the `xtbopt.xyz` file is the most relevant file—this contains the xyz coordinates of the optimized structure. In subsequent sessions, you will be using this as an input to DFT calculations.

4.1 First optimization

Start an interactive job (one core is enough) and optimize your `h2o.xyz` structure. Compare the H-O-H angle before and after the optimization. Try to also optimize the same starting structure as an anion and as a cation (make sure you don't overwrite your results as you run the calculations). Investigate the contents of the resulting files and visualize the resulting structures.

5 Generating structures

5.1 From a SMILES string

The xyz coordinates of many compounds can be conveniently generated directly from ChemDraw. You need to draw the compound of interest, select it, and copy it as SMILES by either selecting Edit-Copy as-SMILES or by `Alt-Ctrl-C` (or `Opt-Cmd-C` on a Mac). You can use our custom script `smi2xyz.sh`, which converts a SMILES string into an initial xyz coordinate guess with Open Babel, and then optimizes it with `xtb`. If it was properly set up, you can just call the script to see the usage guide:

```
smi2xyz.sh
```

For example, to generate the coordinates for benzene, you need to call `smi2xyz.sh` on the SMILES string for benzene (*between quotation marks*):

```
smi2xyz.sh "c1ccccc1"
```

This will run the script, print the progress in the terminal, and generate just the xyz file called "dummy.xyz". If you want to have it named "benzene.xyz" instead, you can type:

```
smi2xyz.sh -o benzene "c1ccccc1"
```

There are also options to specify the charge and the multiplicity of the system, to change the optimization criteria, toggle on/off printing in the terminal (verbose), and to keep/remove all the intermediate files. You have to type the options first with the respective flags and values, and then enter the SMILES string in quotation marks.

For generating coordinates of non-covalent complexes of two or more compounds, you can simultaneously select the separate compounds in ChemDraw. The SMILES string of those compounds will then be copied together, separated by a dot ".". If you try to convert this string into a structure, Open Babel will try to place the compounds at some relative distance and orientation in one xyz—sometimes it works and sometimes it fails. If it fails, you have to generate the coordinates separately, combine them manually in a molecular editor, and reoptimize with `xtb`. Note that you can also copy the SMILES string from Wikipedia articles, if provided, or any other source.

> **Note:** The SMILES string is not generally unique, meaning that one compound can be represented by different SMILES strings. The SMILES string that ChemDraw generates from the drawing depends on various mystical factors, including the order you drew the structure in, as well as whether the atoms are implicit or explicit (e.g., whether a C-H bond is shown in the drawing). For this reason, two apparently equivalent drawings don't always work equivalently well.

5.2 Avogadro

If you need to prepare a small number of complicated structures, Avogadro can be a convenient tool to draw and pre-optimize the structures. Avogadro recently released Avogadro 2 – this is an enhanced but not yet fully functional version. We will tell you how to work with either version.

Drawing is activated by clicking on the pen icon on the upper toolbar (shown in blue in the graphic below). The element and bond order can then be selected in the auxiliary toolbar, which appears after selecting the drawing tool. The structure can be rotated and moved around with the respective tool (shown in green), and the atoms can be dragged around with the manipulation tool (shown in purple). After preparing the structure, the coordinates can be optimized in Avogadro 1 by clicking on the optimization tool button (shown in red) and then clicking "Start" in the respective toolbar. Once the structure looks sufficiently reasonable (usually within a few seconds), the optimization can be stopped. In Avogadro 2 optimization is performed by selecting `Extensions – Open Babel – Optimize Geometry`.

The 3D coordinates can be saved as an xyz file by `File – Save – Save as type: XYZ` (in Avogadro 1) or `File – Export – Molecule` and selecting the xyz file format (in Avogadro 2). In Avogadro 1, you can also copy the coordinates by selecting the molecule (`Ctrl-A`) and copying it (`Ctrl-C`). The coordinates can then be pasted into any text editor.

Avogadro 1

Avogadro 2
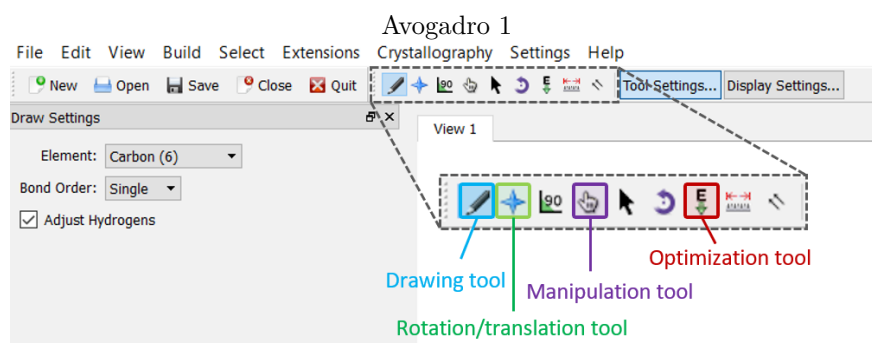
5.3 Chemcraft

Chemcraft also features a few molecule-building tools, which can be accessed by `Edit – Add Atom` or `Edit – Add Fragment`. By selecting the element or fragment from the library, and then clicking in the drawing area, molecules can be built. This is usually not very efficient, unless you intend to modify the structures provided in the Chemcraft database. Common custom structural motifs (e.g., ligands for organometallics) can be added to the personal Chemcraft database by `Edit – Add selected fragments to database`.

To modify the structure, you can use the direct atom or group manipulation tools that you can activate as shown below:

Selection mode active:

Editing mode active:
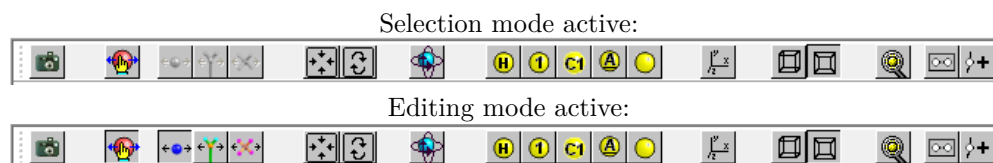
You can either drag atoms (blue icon), groups (green icon), or molecules (pink icon). In selection mode, you can choose a number of atoms and the respective geometric property defined by these atoms will appear in the lower bar of the window:

Select two atoms for a bond length, three for an angle and four for a dihedral angle. You can edit the value in the cell and press enter to effect that change in the geometry.

You can export the edited xyz coordinates by clicking on the "Coord" section, then selecting `Coords format - XMOL Format(Symbol X Y Z)` and then pressing the "Copy" button. You can then paste the coordinates into a text file. Alternatively, you can save the structure as an xyz file by `File - Save as - Save structure`.

5.4 Final notes

The best method for structure generation depends on the task at hand (as well as personal preferences), and occasionally a combination of several can be efficient (e.g., generating a structure with `smi2xyz.sh` and fixing the stereocenters with Avogadro, if necessary). For generating coordinates for a large set of molecules, `smi2xyz.sh` is usually the most efficient method. For any other method, it is still usually useful to optimize the initial structure with `xtb` before proceeding to DFT.

6 Exemplary exercises

> Don't forget to activate the ICC `conda` environment:
>
> `conda activate icc-2023`
>
> This is required to use `smi2xyz.sh` and `xtb`.

We have provided some exemplary cases for you to practice. When you use `xtb` or `smi2xyz.sh`, make sure to make new directories or rename the resulting files so you don't overwrite your results. Possible solutions for the trickier compounds with `smi2xyz.sh` are provided in Section 7. Remember that the result may depend on how you prepared the ChemDraw drawing, so what worked for us might not directly work for you and *vice versa*.
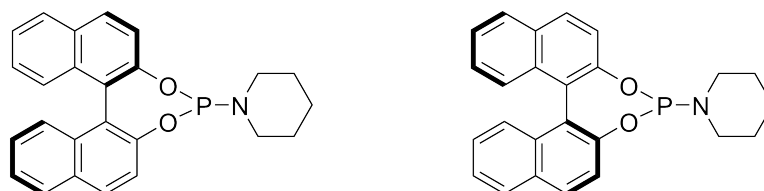
6.1 Simple compounds

Let us start with something simple. Try to generate the 3D structures of the compounds shown below. Note that the SMILES conversion can handle common abbreviations (Ph, *t*Bu, etc), i.e., you don't have to draw everything out.
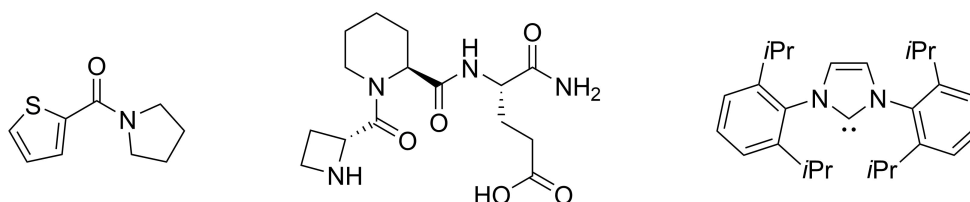
6.2 Stereoisomers

Try to generate the two enantiomers of this phosphoramidite ligand:

Hint: SMILES strings do not contain information about atropoisomers, so the two structures cannot be distinguished, and `smi2xyz.sh` can only generate one of the enantiomers. You can use another option to get the structure or hack your way with `smi2xyz.sh`. For example, you can draw chemically unreasonable bonds to force a specific geometry (see Section 7 for inspiration). You just have to make sure you have the right number of atoms in the right place, and `xtb` will often take care of the rest at the end.

6.3 Ionic compounds

To generate charged structures, you need to add the "-c" flag to the command. For example:

```
smi2xyz.sh -c 1 "SMILES"
```

to generate a cation. You don't necessarily have to include the charge in the ChemDraw drawing—it is only used to generate the initial guess; the charge matters when `xtb` tries to optimize the structure.

Generate the structure of the following acyltrifluoroborate (no charge intentionally shown):
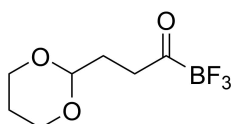
To add a counter-ion, you can often just covalently add it to the anion. For example, to add potassium to the structure above, you can draw:
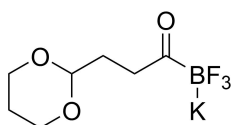
If this does not work for some reason, or you do not want to specify the location, you can also add it as a separate unit. In this case, Open Babel will place it "somewhere" and `xtb` will optimize from that guess. For example, you can achieve the same result as above with:
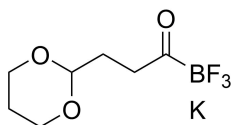

Can you see what the main difference in the SMILES string is for the two options of making the ion-pair?[2]

> **Note:** The outcome of two sequential SMILES to xyz conversions may not be identical for non-covalent interactions. This indicates that you probably have several conformations you should consider in your calculations.

6.4 Stereochemistry in fused rings

Disclaimer: This task can be quite annoying at first if you want to do it with `smi2xyz.sh`. You will need to find which part of the molecule is causing you trouble and then find a work-around. The possible benefit of doing it with `smi2xyz.sh` is that once you figure out the solution, you can quickly derivatize your compound of interest. If you only want one such structure, it is easier to use Avogadro.

Try to find a way to generate the structure of a complex hexacyclic natural product (–)-mitrephorone A:

---
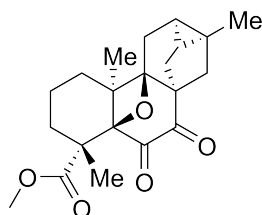
[2] It's the dot for the latter case, indicating that the SMILES string is for two non-bonded structures.

6.5 Organometallics

Most organometallic complexes work fine with `smi2xyz.sh`, but problems can arise with bivalent coordination complexes or π-complexes (and possibly some other cases). There are two fairly simple workarounds:

1. You can replace the metal with a main-group element that has the same valency and which you do not already have in the compound. After `smi2xyz.sh`, you can simply change the main group element symbol in the xyz file to the wanted metal and reoptimize with `xtb`. This can be quite efficient if you want to go through many structures with this problem (requires the least drawing).

2. Try to work your way around the issue by drawing the structure differently. For example, a Ni-ethylene π-complex can be drawn as:
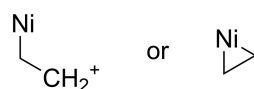

The charge does not matter, it is just there to define two hydrogens on that carbon—therefore, the molecular formula is the same as for Ni-ethylene, although we have not drawn it explicitly as such. Alternatively, you can draw a Ni-cyclopropane. Submitting either of the drawings to `smi2xyz.sh` then returns the Ni-ethylene complex.

Try to generate the structure of the following organometallic systems:
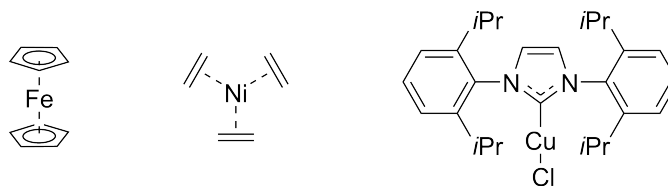
If you succeeded, you can try to decorate the ferrocene with some phosphines to build the dppf ligand:
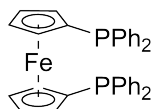
7 `smi2xyz.sh` solutions

This section provides some options for generating the trickier structures in Section 6.

7.2 Stereoisomers

To solve the problem with stereoisomers which have identical SMILES strings, you have to make the SMILES strings non-identical. One way is to draw the compounds in a "nonsense" fashion, and hope that in the end `xtb` will resolve the structure to a physically meaningful one. In this case, you can introduce a bond in one position with the stereochemical information to force the geometry, and charges in adjacent positions to keep the number of atoms the same as in the desired structure (note that the total charge of the drawing does not matter in the calculation, it's just to easily control to the number of atoms). By changing the stereochemistry of the "nonsense" bond, you can access both atropoisomers of the ligand. This dirty trick enables you to rapidly generate many different phosphoramidites.
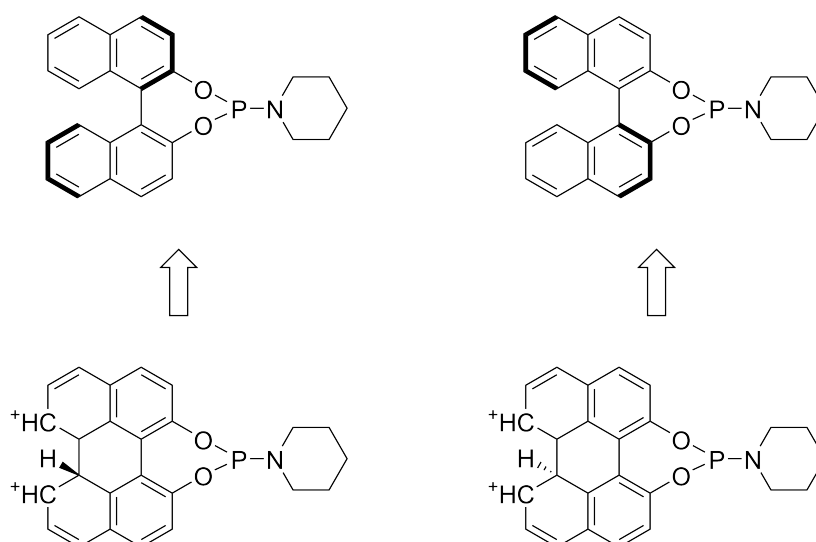
7.4 Stereochemistry in fused rings

Complicated multiring systems don't always offer elegant solutions for SMILES to xyz generation. The best strategy we have is to just test which moieties of the structure are causing you trouble (usually the `smi2xyz.sh` just crashes) and then again use "nonsense" drawing to get around those issues. The most usual offenders are stereocenters at the fused bonds and strained ring systems. For (–)-mitrephorone A, the drawing shown below corresponds to a SMILES string that converges to the correct structure. The offenders in this case are the oxetane moiety (for which removing explicit stereochemical information still luckily leads to the correct isomer), and the cyclopropane moiety. Note again that the total charge of the drawing does not matter for our script, and the charges just enable convenient handling of the number of H atoms.
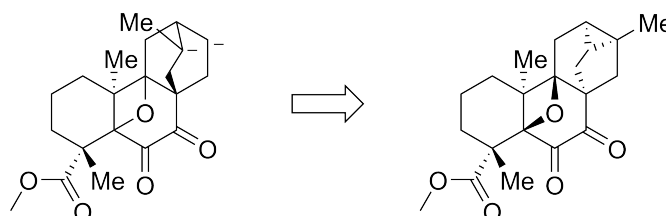
7.5 Organometallics

Ferrocene can be generated by drawing out all the Fe-C bonds explicitly. The Ni π-complex can be generated by drawing three metallacyclopropanes. Odds are that [CuIPr]Cl worked fine for you as drawn in the exercise. In case it didn't, it can be generated by having the chloride as a separate unit (so the SMILES string will have a dot in it), and letting `xtb` optimize the position of the chloride. Alternatively, you can substitute Cu for P and after running `smi2xyz.sh`, you can change P back to Cu in the xyz file and reoptimize with `xtb`.