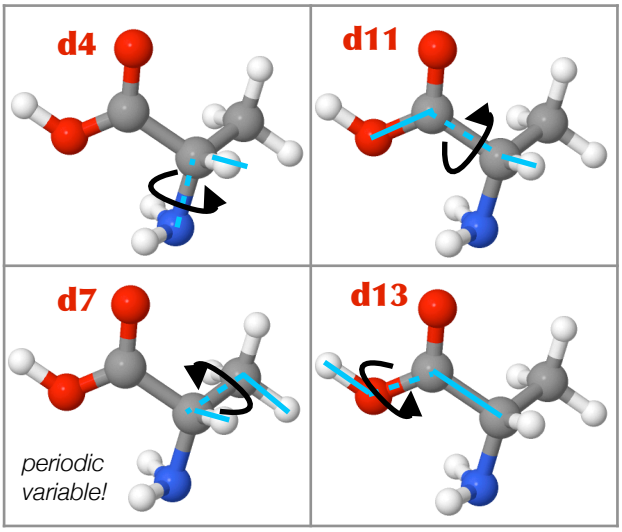
Machine Learning the Structure of Alanine
Tutorial script

In this Tutorial, you will learn how to build and use machine learning models, namely Bayesian Optimization (BO). BO iteratively builds a probabilistic model for a an unknown function, with the aim to find the global minimum. In this example, our unknown function is the potential energy landscape of the alanine molecule. The parameters are the 4 torsional (dihedral angles) that define different alanine conformers. Our function is 4-dimensional.

Your objective: Use BO to find the lowest energy of the alanine conformer with respect to the torsional angles.



Necessary software:

- Git Bash terminal window - open immediately and check you can type there.
- Bayesian Optimisation Structure Search (BOSS) tool. Type “boss” into terminal to check it is there.
- Jmol software for viewing molecular movies - open from Windows to check it is there.
- Notepad software for opening text files.

Part I: Alanine 2D search

1. Open the Git Bash terminal window.
2. Enter the Desktop and the Tutorial directory, list the contents.

```
#> cd Desktop
#> cd BOSS_Tutorial
#> ll
```

3. We will begin by performing a 2-dimensional structure search for alanine. We will vary torsional angles d4 and d13. Enter the Alanine_2D directory and list the contents.

```
#> cd Alanine_2D
#> ll
```

4. Open the BOSS input file `boss_2D.in` with the Notepad. We will query two angles around 360°, taking 5 initial steps and then 40 BO steps, note how this is defined.
5. Open another Bash terminal window so that they are side-by-side and enter the same directory (Alanine_2D). In the first, you will execute the BOSS simulation. In the second you will follow the output as the simulation is running.
6. In the first terminal start the 2D BOSS simulation by typing the following:

```
#> boss op boss_2D.in
```

7. In the second terminal, start watching the BOSS output by typing:

```
#> tail -f boss.out
```

8. Can you see the iterations being written out? The simulation should take 5 min. At the end, you will see the “Have a nice day” message. The first terminal is now free for typing.

9. When the simulation is finished, we will carry out some analysis. Open the boss.out file with Notepad. Scroll down to BO iteration 50, and find the words `Global minimum prediction` (column 3). What is the lowest energy found? Write it down below to 3 decimal points.

E_{2D} =

10. Using Windows, open the `postprocessing` directory in `Alanine_2D`, then `graphs_models`. Switch from list to image view. You will see plots of 2D energy landscapes for each iteration (ignore “uncert” images). Do the landscapes look converged at 50 iterations?

11. Open the energy map at iteration 50. What is the lowest point found on the map (hint: see bottom of colormap scale). Does it match the value from the file?

12. Let’s view the movie of the structures BOSS sampled. Open the Jmol program from the Windows. From the window of the Alanine_2D directory, drag the `movie.xyz` and drop it into the Jmol window.

13. You should see the alanine molecule on a black background, rotate it with the mouse so you can see the conformations well. On the Jmol toolbar, select “Tools”, then “AtomSetChooser”, and a new window will open. Click on the first “E=…” row in the white area - this is the energy of the first structure sampled. Now you can drag the “Select” arrow left and right to view the entire movie - the corresponding energies appear in the “Info” field.

14. Can you find a structure with energy close to the lowest predicted? What is happening with the two functional groups here?

Part II: Alanine 4D search

1. Exit the 2D directory and enter the 4D one, list the contents.

```
#> cd ..
#> cd Alanine_4D
#> ll
```

2. Open the BOSS input file `boss_4D.in` with the Notepad. We will now query four angles around 360°, taking 5 initial steps and then 70 BO steps, note how this is defined.

3. Open another Bash terminal window so that they are side-by-side and enter the same directory (Alanine_4D). In the first, you will execute the BOSS simulation. In the second you will follow the output as the simulation is running.

4. In the first terminal start the 4D BOSS simulation by typing the following:

```
#> boss op boss_4D.in
```

5. In the second terminal, start watching the BOSS output by typing:

```
#> tail -f boss.out
```

6. The simulation should take 15 min. Leave it to run.

7. After it is finished, open the boss.out file with Notepad. Check `Global minimum prediction` (column 5) for the last BO iteration. Write down the lowest energy below to 3 decimal points.

E_{4D} =

8. Has the 4D simulation found a lower energy minimum than the 2D one?