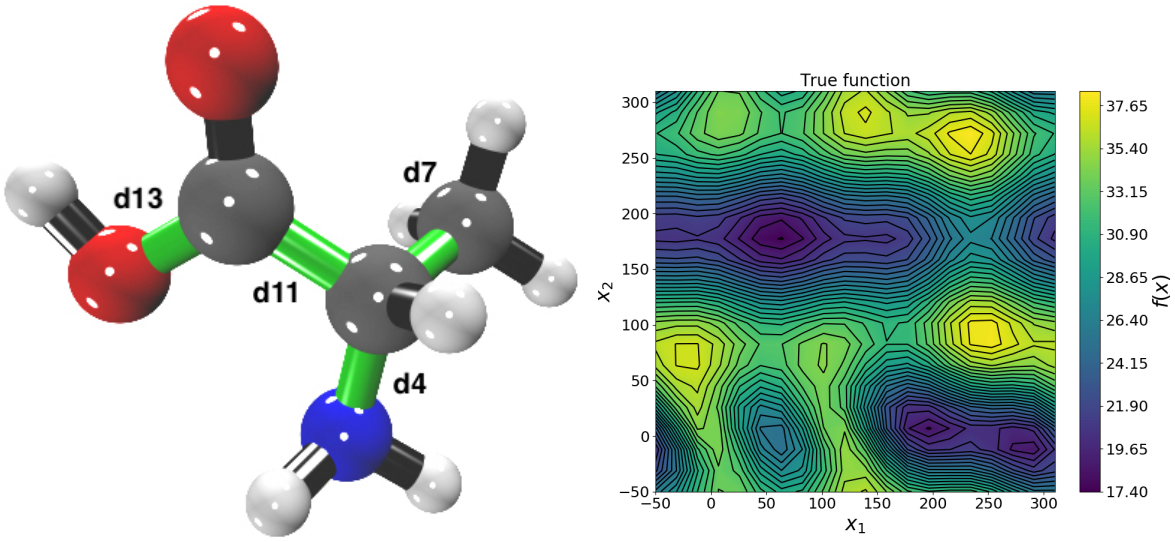
BOSS Demo 2
Periodic conformer search

1 Problem

Many organic molecules can take on different shapes depending on how their atomic bonds are directed and functional groups positioned. The lowest energy conformers are particularly important, since they are often the forms the molecule takes in nature. For the amino acid alanine (see reference structure in figure 1a) all the relevant conformers can be obtained by changing the dihedral angles. We access them by switching to internal coordinate representation. The dihedrals correspond to rotating the bonds marked with green in figure 1a.



(a) Alanine reference structure

(b) True 2D PES (d4-d13)

Figure 1

The total set of simulation variables are the four dihedral angles, but we also consider a simplified 2D subproblem, where only two dihedrals are allowed to vary (while the others are fixed to their value in the reference structure in figure 1). These two are the bond leading to the amino group (d4) and the one leading to the hydroxyl group (d13). The true potential energy surface (PES) spanned by these two variables was calculated in a grid and is shown in figure 1b. Alongside to the 2D problem, we consider the full 4D conformer search problem, to find most stable alanine conformer.

2 Run instructions

This demo features separate optimization and post-processing parts for both the restricted 2D problem and full 4D problem. There are two prerequisites for user function part: 1) AMBER forcefield for simulations and 2) openbabel for converting between internal and cartesian coordinates. You may load AMBER as a module on your desktops and Taito, and find openbabel from your system or install it. The python user function script writes the requested variable values in a file `energy_calc/variables.in`, calls a shell script to build and simulate the structure, and reads the resulting energy from file `energy_calc/energy.out`. The shell script operates in a directory of its own called `energy_calc`.

Run tasks

1. Study the input files and user function scripts in detail. What are the differences between 2D and 4D cases? How do `userfn_2D.py` and `userfn_4D.py` interact with the machinery that runs the atomic simulations?

2. Familiarize yourself with the shell scripts in `energy_calc` by reading the comments in the script. Note that the scripts do useful data storing operations as well.

3. Starting with the input files `in_2D` and `in_4D` run optimization only, and store the produced results in a directory of their own (e.g. `./move.sh 2D_results`).

4. Decide how you wish to post-process the results and modify the restart files accordingly (at least remove the user function keyword). Then run `boss p modified.rst boss.out`.

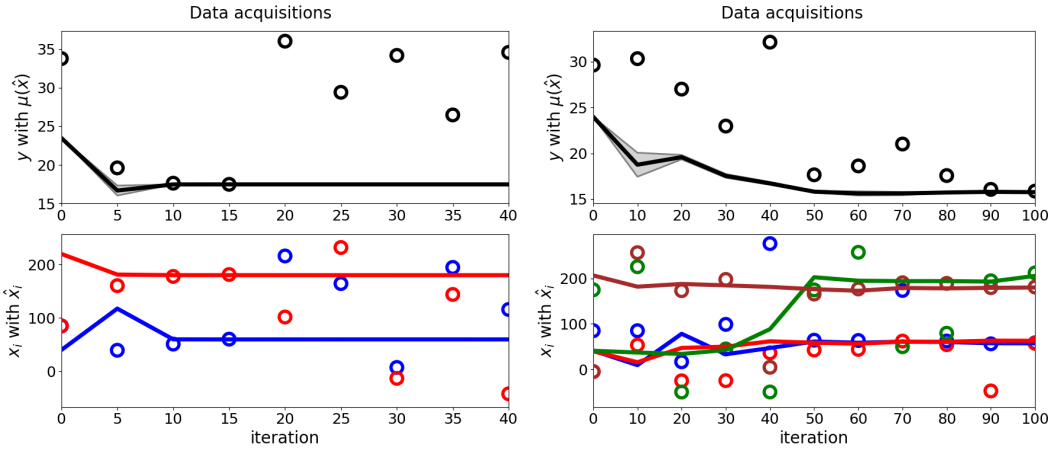
5. Analyze the results. Has the global minimum prediction converged in both cases? At which iteration? How about the entire model (or 2D slice of the model in 4D)? Compare to the true PES shown in figure 1b.

6. Visualize the structures that were acquired by BOSS (using `movie.xyz`). What does the most stable alanine conformer look like? Hint: repeat the acquisition process manually by writing the converged global minimum prediction variable values into `energy_calc/variable.in` and calling the shell script.

7. Compare the results you obtain with those shown in section 3.

3 Results and Analysis

The main results of the 2D BOSS search can be seen in figures 2a and 3a. The data acquisitions figure shows that, after iteration 15, the global minimum was no longer targeted by acquisitions, nor did it change its location. This is typical behavior of the acquisition functions BOSS uses, and it indicates the global minimum prediction is reliable. The quality of the model map can be checked against the pre-computed true map (figure 1b). The GP model of the PES closely resembles the true one. All aspects of the landscape match and the energy scale is similar.

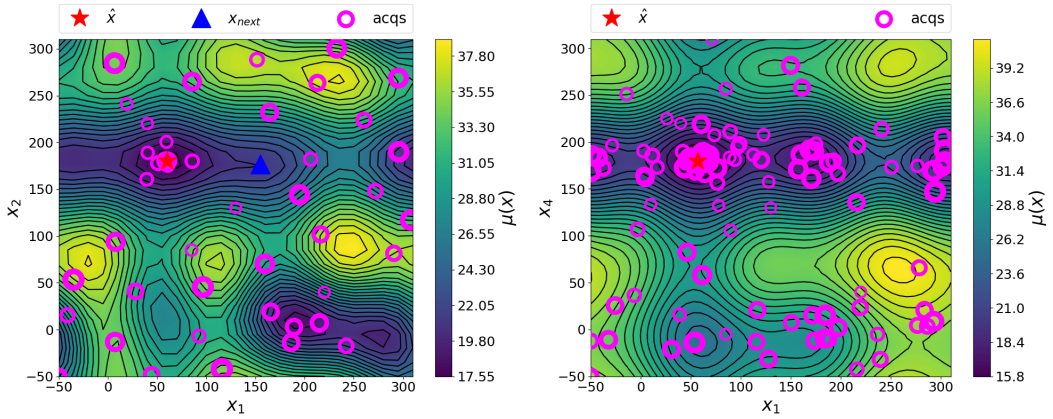


(a) 2D

(b) 4D

Figure 2: Data acquisition locations and values compared to current global minimum prediction in 2D (a) and 4D (b) runs.

The data acquisitions of the full 4D conformer search in figure 2b show a much smoother and slower convergence of global minimum prediction. There is a big change in one of the variables value (green line) for global minimum prediction at about iteration 50, although there is no corresponding drop in the predicted energy. Towards the end of the run, global minimum prediction seems to stop changing. Nevertheless one might want to run more iterations yet, to become more confident about the convergence. Looking at the 2D slice of the GP model (figure 3b) cut in the same plane as the preceding 2D search was done, we can see that this projection of the 4D global minimum corresponds to the 2D-only case. Note that the 2D projection of the 4D map should not necessarily match exactly since the values of the remaining two variables may be different. Because alanine is a well studied molecule, we can compare the 4D global minimum prediction in figure 4 to literature, which confirms that we did indeed find the most stable conformer.



(a) 2D - iteration 40

(b) 4D - iteration 100

Figure 3: Final GP models from 2D (a) and 4D (b) BOSS searches. The 2D slice taken from the 4D GP model is cut from the same plane (variables d4 and d13) that was used in the 2D.

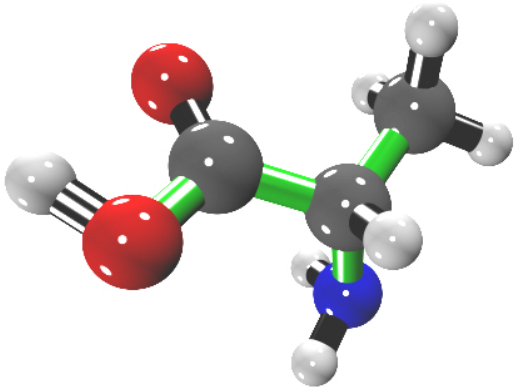


Figure 4: Most stable alanine conformer found in BOSS 4D search.