ME 5329 Project 1

Nucleotide Dimer Calculation using CREST and ORCA

September 4, 2025

Dr. Aquino Fall 2025

Due: September 11, 2025

Objectives:

- Learn and understand what xTB is and the GFN2-xTB method by Grimme.
- Learn about CREST and what a metadynamics (MTD) simulation is.
- Understand the different levels of theory being used in this assignment and why our workflow is this way.
- Learn what a monomer, dimer, trimer etc. structure is and how to make one using `molecule_lib`.
- Discover the effects of a dimer structure in a nucleotide.
- Learn about molecular orbitals and discover how HOMO-LUMO orbitals change during stacking.

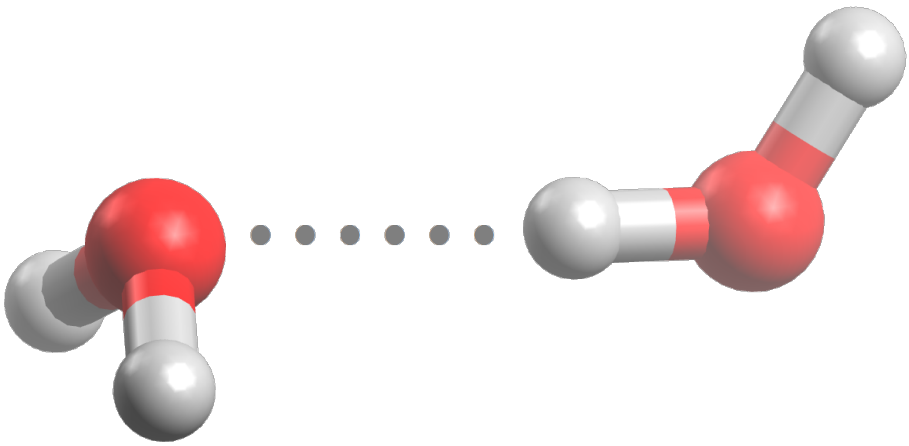

Figure 1: Example of a dimer structure for water (H_2O).

Task 1: Dimer Structure Creation

1. Perform the calculations and analyses below for each of the following DNA nucleotides:
    - Adenine
    - Cytosine
    - Guanine
    - Thymine
2. Create an initial DNA nucleotide molecule monomer structure from the SMILE string or your preferred method and save as a .xyz file.
3. Using xTB, optimize this monomer structure using the GFN2-xTB method.
    - You can learn more about xTB here: xTB Tutorial
    - CREST uses xTB, so you only need to create/activate your CREST conda environment. The command to create the conda environment is "conda create -n crest conda-forge::crest" and to activate the conda environment is "conda activate crest".
    - You may also use Jeremy's conda environment from instructions given in class.
    - To optimize a structure using xTB, you may run this command in the terminal for example, "`xtb NUCLEOTIDE.xyz --opt tight > xtb.out`"
4. Using `molecule_lib`, load the xTB optimized .xyz file using `mlb.read_xyz()` and save as a `XYZMolecule` instance.
5. On the original `XYZMolecule` instance, use `XYZMolecule.add_coords()` and also use the original `XYZMolecule` instance for the "molecule" argument.
    - The `XYZMolecule.add_coords()` uses `XYZMolecule.move()` and `XYZMolecule.manipulate()` if you want to learn more about how it works. Read the documentation to learn more if you are interested.
    - If you notice the dimer is not perfectly stacked, that means the original structure is not perfectly orthogonal in a specific axis plane. You can use the `XYZMolecule.rotate()` method to fix this issue.

6. Below shows an example Python code of how this is acheived using `molecule_lib` to create a nucleotide dimer that is 4 Å seperated in the z-axis direction. Between 1.5 to 3 Å is a good distance to start for a dimer structure.

```
import molecule_lib as mlb
monomer_mol = mlb.read_xyz("NUCLEOTIDE.xyz")
dimer_mol = monomer_mol.add_coords(molecule=monomer_mol,
                    absorbent_reference="Bottom",
                    surface_reference="Top",
                    dist=4,
                    axis="z")
dimer_mol.to_xyz("NUCLEOTIDE_DIMER.xyz")
```

7. There are other ways of creating this dimer structure. You can also use Avogadro, MS Excel, Python or by hand in a text file to do it also. You may use a different method as long as it is explained in the report.

Task 2: Use CREST to Optimize the Nucleotide Dimer

1. Familiarize yourself with the CREST tutorial linked below:
   - CREST Tutorial
2. Obtain the SLURM file from the link below:
   - CREST SLURM File
3. Ensure your input geometry file for CREST is named "input_geometry.xyz".
4. Submit the CREST job using SLURM. CREST will print its output to the "crest.out" file.
5. CREST if finished when it says *"CREST completed normally"* at the bottom of the "crest.out" file and files such as "crest_best.xyz" and "crest_conformers.xyz" are created. The lowest energy conformer is located in the "crest_best.xyz" file and this will be the geometry we will use as an initial structure in the higher level of theory calculation, ORCA.

Task 3: ORCA Optimization and Frequency

1. For the monomer and dimer structure (which is previously found from xTB and CREST), do an ORCA geometry optimization using KS-DFT functional B3LYP and basis set TZVP and calculate the vibrational frequencies of the optimized structure using ORCA. The charge of each system is 0 and the multiplicity of this system is 1.
   - The ORCA input line should look like this: *"! B3LYP TZVP OPT FREQ"*

2. Report the first 10 frequencies for each structure.

- If a negative frequency is found for a structure, what could that possibly mean?

3. Report the HOMO-LUMO gap in eV for each structure and the effect of the dimer. The equation is below.

- $E_{gap} = E_{LUMO} - E_{HOMO}$

4. In a table, report the interaction energy in kcal/mol and kJ/mol of the dimer using the equation below for the two levels of theory we used, xTB and ORCA.

- $\Delta E_{interaction} = E_{dimer} - 2 \cdot E_{monomer}$
- Look in the crest.out file to find E_{dimer} for the xTB level of theory $\Delta E_{interaction}$ calculation.

5. Report the visualized HOMO-LUMO orbitals for the monomer and dimer structure. Discuss the differences between each structure.

- To do this, you first need to generate the molden.input file by using this command in the ORCA calculation directory, "orca_2mkl orca -molden".
- The molden.input file can be read by MOLDEN and JMOL. To visualize the HOMO and LUMO orbitals in JMOL, you need to open up the JMOL console by clicking on **File** then **Console**. The JMOL console command to visualize the HOMO is "mo homo" and likewise for the LUMO. You may also view each molecular orbital by using the command "mo [int]" where the integer is the numbered molecular orbital you want to view.
- Some other settings you may want to try in JMOL console to make the viewing of molecular orbitals better are listed below.
  - "background white"
  - "mo fill translucent 0.5"
  - "mo colour blue red"
  - "mo cutoff 0.02"
  - "mo nomesh"

6. Compare the dimer structure found from CREST and ORCA. Identify if there are any major differences. Visualize the two optimized dimer structures and measure the distance between them, show these in the report. Compare the distance between the molecules found to the initial dimer geometry generated.

Submission Requirements:

Please include the scripts you used to generate the structures or information on where the structures were found, paths to where the calculations were ran on the HPC (a parent directory for this specific project works also) and your post processing scripts that generated the graphs and results in the Canvas Assignment submission. Including this information is the student's way of showing work in this class. If the information requested above is not provided, the instructor will assume plagaraism or collusion occured and respond accordingly.

The only accepted submission format for the report is a Microsoft Word document.

Please submit it under the **'Project 1'** assignment in Canvas.