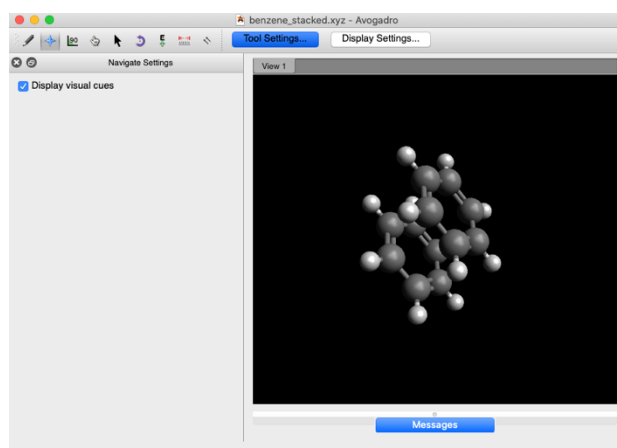


In *Lab 2*, we will geometry optimize more complex structures. We will also help you set up the command line on your computer if you run into any trouble. You will need to complete a worksheet with questions about key results from today's lab. While completing this lab, record your response to each **worksheet prompt** as you encounter them. We ask you to submit the **entire worksheet by 11:59 PM EDT 09/21/23 (Thu)**.

In today's lab, we will work on both computing binding energies (part A) and then conformational (or *conformer*) searches (parts B and C) using some standard force fields. We will also be going over some basics of using the command line and accessing ACCESS (part D). These tools will be critical for future labs.

A. Build two benzene molecules and study their interactions. Stacked “sandwich” benzene molecules have been studied extensively using *ab initio* methods to understand how well methods capture π -stacking interactions. There is some debate whether the sandwich configuration (one benzene stacked on top of the other like they are mirror images) is a minimum energy configuration or a saddle point. Regardless, interaction energies from the most accurate *ab initio* approaches, i.e., coupled cluster theory with singles, doubles and perturbative triples – CCSD(T) – extrapolated to the complete basis set limit (a topic that we'll cover later this term), are favorable. Results from CCSD(T) suggest a favorable interaction on the order of **7.32 kJ/mol** with inter-benzene C-C distances of around **3.90 Å**. Experiments suggest an interaction energy of 8-12 kJ/mol as well, though generally the T-shaped dimer is preferred experimentally.

You could build the benzene dimer by hand (see section C in Lab 1), but we have constructed this initial structure for you. (Download **benzene_stacked.xyz** from Canvas). This copy of the initial coordinates aligns the plane of the two benzene rings perpendicular to the z-axis, which will also make it easier to use constraints on their Cartesian coordinates during the optimization.



Window 1. The initial structure of the stacked benzene dimer.

- 1) Download the initial structure of the stacked benzene from Canvas at **Modules/Lab 2: Geometry optimizations** named **benzene_stacked.xyz**. Open the **.xyz** file in Avogadro. Your screen should resemble the one shown in *Window 1*.

(Optional Windows Note) Open Avogadro first, and open the **.xyz** file from inside Avogadro, otherwise force fields may not work.

- 2) Use the *Measuring* tool to mark the distance of two carbon atoms that are positioned directly over each other.
- 3) Use the *AutoOptimization* tool to optimize the structure with the **MMFF94** force field.

Worksheet prompt: Record the final energy (in kJ/mol) for the optimized structure as well as the C-C distance (in Å).

Note: The potential energy surface here is very flat! Thus, it will take a very long time to reach full convergence as defined by the Avogadro threshold (10^{-7} kJ/mol on energy). We recommend that you manually stop the optimization earlier and/or increase the threshold for convergence in **Extensions>Molecular Mechanics> Setup Force field**.

Note: You will notice that the benzene rings will slide relative to each other. Stop the optimization right before (or right as) the sliding starts.

- a. We could address this by adding constraints to the optimization. We want to avoid translation of the two benzene molecules with respect to each other and all rotation. Let's start by fixing an atom.

Worksheet prompt: What additional constraints might help with studying the benzene-benzene stacked interaction?

- b. Bring up the constraint window by going to **Extensions > Molecular Mechanics > Constraints**. Select the "Fix Atom" type of constraint and apply it to atom 1 (C atom), which is in the first benzene.

To find atom indices:

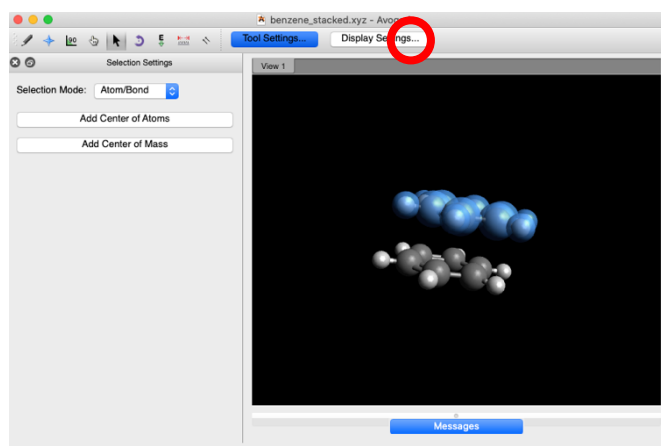
Click on *Display Settings...*

- i. Check "Label" in the *Display Settings* dialog box
 - ii. Click the wrench to the right of "Label"
 - iii. Select Atom Number from the "Text" dropdown box.
 - iv. Hide Display Settings again.
 - v. Now you can see which atom numbers correspond to each atom.
-

- 4) Now change the force field to GAFF and re-optimize the structure.

Worksheet prompt: Record the **GAFF** energy and C-C bond distance.

- 5) Separately, find the energy of a single benzene molecule using each of the force fields (i.e., MMFF94 and GAFF). This can be accomplished by building a benzene molecule by hand (like we did in lab 1) **or** by deleting one benzene from the dimer structure. You can delete atoms by selecting them with the selection tool (the mouse pointer) and using the **Del/Delete** key.



Window 2. The selection tool is circled in red. The selected atoms are highlighted in blue.

Hint for SA: For reporting energetic differences, the tenths place in kcal/mol is sufficient, and the hundredths place in kJ/mol is sufficient. The hundredths place for geometric measurements (in Å or degree (°), where relevant) is also sufficient.

Worksheet prompt: Fill in the table for part A of your worksheet that summarizes some of the properties of two stacked benzene molecules.

Worksheet prompt: How does the **GAFF** binding energy compare to the CCSD(T) reference? Is it an overestimate or underestimate? What about the separation distance? Is it an overestimate or underestimate?

Worksheet prompt: How does the **MMFF94** binding energy compare to the CCSD(T) reference? Is it an overestimate or underestimate? What about the separation distance? Is it an overestimate or underestimate?

B. More chemically diverse molecules and conformer searching. As we just observed for benzene in section A, complexes can prefer multiple orientations. In addition to this challenge, for all but the simplest of molecules, there are also multiple *conformers* that are often close in energy, which correspond to different orientations of rotatable groups within a molecule. Researchers often build and simulate these different conformers to compare their energies. In this next section, we'll use the built-in force fields in Avogadro to compare conformer energies.

In particular, we'll look at several molecules to identify whether a gauche (~60° tilted) or anti (180°) dihedral orientation is preferred for substituted alkanes as well as how this is influenced by through-space hydrogen bonding interactions in larger molecules. In select cases, a gauche structure is the low energy structure in the gas phase due to electronic effects, while more commonly an anti structure is the low energy structure because it minimizes steric repulsion. Thus, comparing these conformer energies is also a sensitive test of the molecular modeling method.

The main molecules you'll be working with today are butane and 1,2-ethanediol. Typically, the anti conformer is more stable than the gauche conformer, but 1,2-ethanediol is a special case

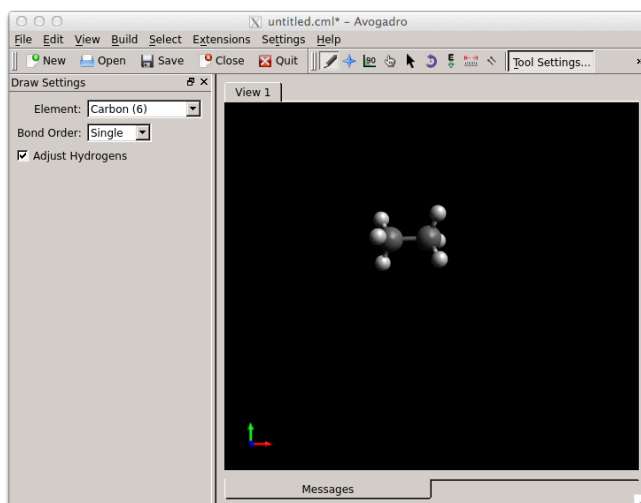
where the opposite is true due to strong electrostatic attraction from an intramolecular hydrogen bond.

Goal: We would like to see if the force field used for the conformer search can correctly predict the gauche effect in these molecules.

For the sake of building all of the molecules, it makes sense to start by drawing ethane, adding hydrogens, and then optimizing the geometry with the *AutoOptimization tool*. Butane and 1,2-ethanediol can be constructed by substituting hydrogens for methyl or hydroxyl groups.

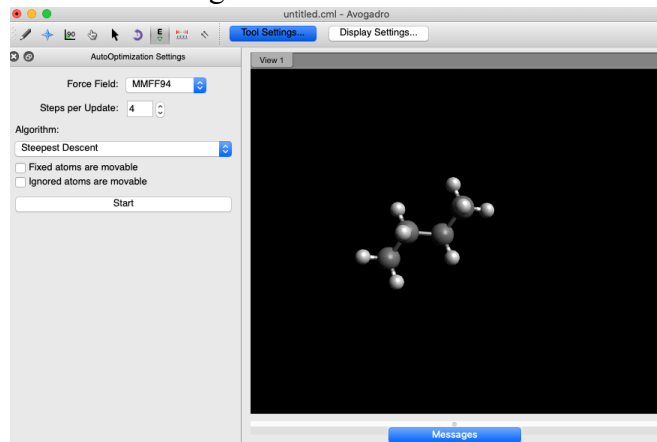
Now, we'll carry out the conformer search.

- 1) Begin by constructing and optimizing the structure of an ethane molecule, as we did in lab 1. When you are finished, your screen should look like *Window 3*. We'll start with the **MMFF94** for field, though we'll eventually want to do a conformer search with both **MMFF94** and **GAFF**.



Window 3. A simple ethane molecule.

- 2) Generate butane by changing one hydrogen to a methyl group on each of the two carbons and re-optimizing the structure. You can alternatively generate the butane molecule with a SMILES string as we learned in *Lab 1*. The result should look like *Window 4*.



Window 4. A simple butane molecule.

- 3) Now that we have constructed a butane molecule, we are ready to carry out conformer search. Go to **Extensions>Molecular Mechanics>Conformer Search...**
- 4) Use the **Random Rotor Search** and identify 10 conformers. Many of these will be duplicates due to the high symmetry of the butane molecule.
- 5) Let's look at both the energetics and the geometric properties of our conformers. To look at the **energetics**, we go to **View>Properties>Conformer Properties...** and a window should pop up that looks like a table. When you first bring it up, the conformers will not be sorted by energy, but you can sort them by **clicking on the energy tab**. The lower the number, the more stable the conformer is. These tables should resemble *Window 5*.

Conformer	Energy (kJ/mol)
Conformer 1	-21.2518
Conformer 2	-17.9767
Conformer 3	-17.9754
Conformer 4	-21.2518
Conformer 5	-17.9767
Conformer 6	-17.9754
Conformer 7	-21.2518
Conformer 8	-17.9767
Conformer 9	-17.9754
Conformer 10	-21.2518

Conformer	Energy (kJ/mol)
Conformer 1	-21.2518
Conformer 4	-21.2518
Conformer 7	-21.2518
Conformer 10	-21.2518
Conformer 2	-17.9767
Conformer 5	-17.9767
Conformer 8	-17.9767
Conformer 3	-17.9754
Conformer 6	-17.9754
Conformer 9	-17.9754

Window 5. Table of conformer energies (unsorted, left and sorted, right).

- 6) By analyzing the list of conformers by energy, we should notice that there are only **two unique conformers (i.e., local minima) of butane that are found by this search**. Display either conformer in the viewing window by first selecting it in the conformer properties window.

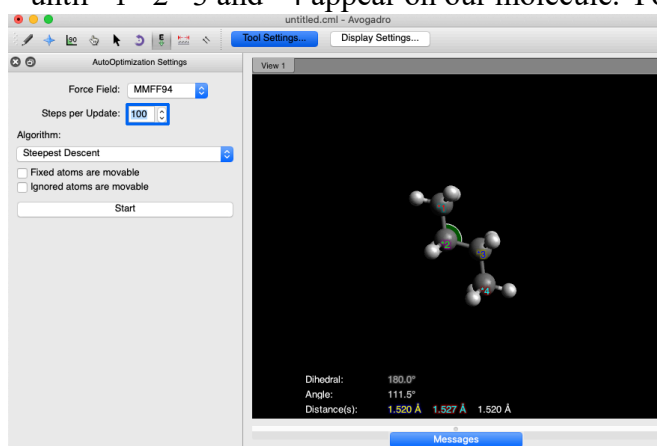
Note: You may further optimize each conformer using the *AutoOptimization* tool.

Alternate Conformer Search. If you have trouble obtaining the conformers as described (**e.g., on a Windows PC**), you can try the following to obtain them with constraints – we constrain the dihedral to a target value (e.g. 60° for gauche or 180° for anti) then we'll release that constraint:

- a) Ensure you are using the **MMFF94** force field in the energy minimization.
- b) Click on *Display Settings...*
 - a. Check "Label" in the *Display Settings* dialog box
 - b. Click the wrench to the right of "Label"
 - c. Select Atom Number from the "Text" dropdown box.
 - d. Hide Display Settings again.
 - e. Now you can see which atom numbers participate in each bond.
- c) Navigate to **Extensions>Molecular Mechanics>Constraints**
 - a. Note the 4 atoms participating in the dihedral of interest shown in *Window 6*.
 - b. Under add constraints, select "Type: Torsion Angle".
 - c. Click *Add*.
 - d. Set the "type" to "torsion angle" and select the 4 atoms that define the dihedral angle shown in *Window 6*.
 - e. Set constraint value, e.g. to 60° for gauche or 180° for trans.
 - f. Click *Add* and then *OK*

- d) Under measure, click on the four atoms to define your dihedral. It should display in the bottom left hand corner.
 - e) Start an **MMFF94** energy minimization. You should see “Num Constraints: 1”.
Note: we need to use **MMFF94** because the constraints work with this force field but do not work with other force fields in Avogadro.
 - f) Once the energy minimization is finished, write down the dihedral at the bottom left corner and the energy at the top left corner. It will not perfectly match your constraint since this constraint is harmonic and the force constant is finite. If you want to, you can add duplicates of the same constraint. The force constants will be additive and you’ll get closer to your target value for the constrained dihedral.
 - g) Go back into **Extensions>Molecular Mechanics>Constraints** to delete the constraint. Re-optimize the structure without the constraint to obtain the closest local minimum.
-

- 7) **Measure the dihedral angle in butane.** Click on the measuring tool icon in the toolbar. We’ll identify the dihedral of butane by clicking sequentially on each of the 4 carbons until *1 *2 *3 and *4 appear on our molecule. Your screen should resemble *Window 6*.



Window 6. Measuring the dihedral angle in butane.

- 8) Measure the dihedral angle of both the anti and gauche conformer of butane using MMFF94. Record these values and save an .xyz file of each conformer, which you will need later for your lab report.
- 9) Select a gauche conformer and re-optimize it with **GAFF**. Do the same for an anti conformer. Record the values for the **GAFF** energies and dihedral angles of each conformer.
- 10) **Repeat steps 1 through 9** for 1,2-ethanediol. 1,2-ethanediol can be generated by changing a hydrogen in ethane to a hydroxyl group instead of a methyl group in step 2. There are actually 10 (!) conformers of 1,2-ethanediol, although most are unlikely to be found by this conformer search due to both the force field and their position on the global potential energy surface. *Interested in more about this? Check out Reiling et al., J. Comput. Chem. (1996) (available on Canvas in the Topic 2 section).* For the 1,2-ethanediol system, there are at least two gauche conformers. In one conformer, the hydrogen atoms are oriented away from each other. In the other conformer, one hydrogen atom points toward the opposing oxygen atom and forms a hydrogen bond. Use

the manipulation tool (the gloved hand) to manually "tug" the hydrogen into the appropriate location.

Hint: The strongest hydrogen bond you should be able to form will be one in which the O-H...O angle is nearly linear. You should aim to form this one because it is the strongest, although you may be able to find others as well.

Hint for SA: Avogadro reports energetics in kJ/mol, while the numbers I will give you in this write-up are often in kcal/mol. If I am asking you to compare to kcal/mol, convert your energy from Avogadro to kcal/mol. $4.184 \text{ kJ/mol} = 1 \text{ kcal/mol}$.

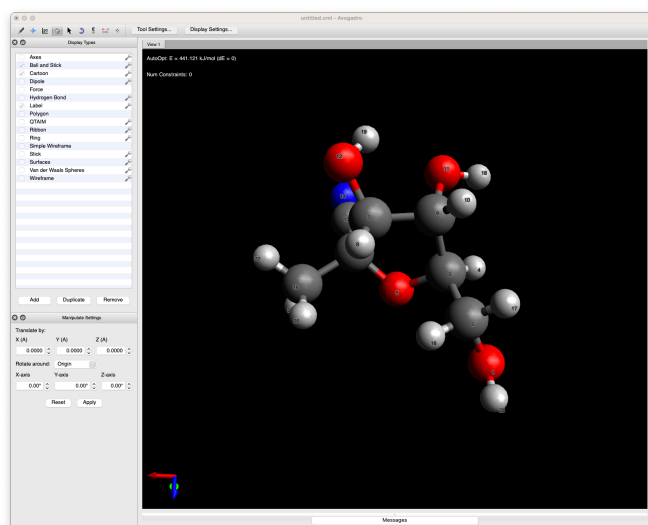
Worksheet prompt: Fill in the table for part C that summarizes the properties of gauche and anti conformers of butane and 1,2-ethanediol.

Worksheet prompt: Based on the table from the previous prompt, why is the gauche conformer lower in energy for 1,2-ethanediol?

Worksheet prompt: How strong is the 1,2-ethanediol hydrogen bond according to MMFF94? According to GAFF?

C. Exploring hydrogen bond preferences in a model component of remdesivir.

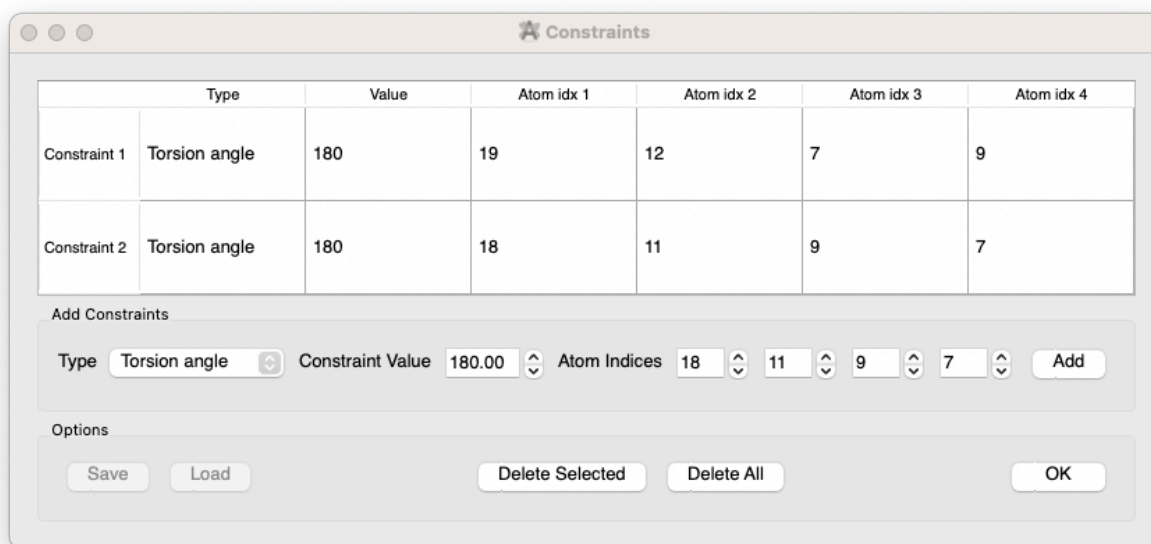
We have extracted the central tetrahydrofuranyl ring in the therapeutic drug remdesivir for you to analyze the hydrogen bonding preferences similar to 1,2-ethanediol but with the added complication of ring strain. Download this .xyz file and load it in Avogadro. Your molecule should look like this:



Window 7. Remdesivir substructure in Avogadro.

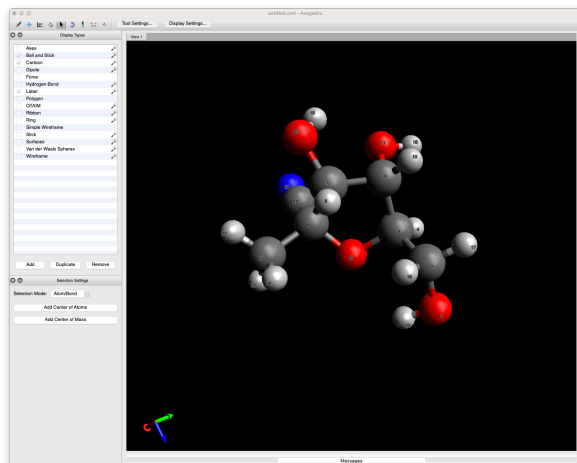
1. Reoptimize this structure with MMFF94, noting in each case the total energy after optimization. If the optimization keeps running for a long time, feel free to stop the optimization once the energy is not changing by more than 1 kJ/mol.

2. Take the hydroxyl that is currently hydrogen bonded to the other hydroxyl on the ring and rotate it away using the mouse selection tool followed by the hand selection tool. Add two constraints to enforce this geometry and run a constrained geometry optimization:
 - a. Turn on *Label* under **DisplayTypes** if you do not already have labels on. Note that you may need to adjust the Label settings by clicking the wrench icon next to Label to display the atom numbers.
 - b) Using the labels which should be 19,12, 7, 9 and 18, 11, 9, 7 set two torsion angle constraints of 180 degrees.
 - c) Re-run the optimization and note the energy with MMFF94.



Window 8. Recommended constraints for the atoms corresponding to the hydroxyl groups in the remdesivir model.

3) There's another hydrogen bond you could form between the oxygen on the tetrahydrofuran ring and the terminal methyl - reorient this one to form a hydrogen bond and rerun the optimization.



Window 9. Example of how the remdesivir model should look after reorienting the second hydrogen bond.

Worksheet prompt: Based on the difference in conformer energies you obtained in 1) and 2), which hydrogen bond is stronger for MMFF94, the strongest one in 1,2-ethanediol or the one in the remdesivir model? What is the dihedral angle for the O-C-C-O angle result from 1)?

Worksheet prompt: Based on the difference in conformer energies you obtained in 2) and 3), which hydrogen bond is stronger for MMFF94, the strongest one in 1,2-ethanediol or this second one in the remdesivir model?

Worksheet prompt: Based on the trends you've seen in hydrogen bonding thus far, if you wanted to strengthen binding between a drug molecule and a protein by 75 kJ/mol, and you knew the protein had up to 8 hydrogen bond acceptors in the active site, how many hydroxyl groups should you add to the druglike molecule assuming they would form intermolecular hydrogen bonds with the protein?

Hint: Examine an intermolecular hydrogen bond by putting two molecules that form a hydrogen bond close together, optimize the structure, then freeze all atoms except the hydrogen, and rotate it away from the hydrogen bond acceptor. You may use two serotonin molecules, a serotonin molecule and the remdesivir model molecule, or whatever else you would like.

D. Accessing and getting familiar with the Linux command line. For the majority of remaining labs in this class, we will be working at the Linux command line. This tool is essential for navigating the world of computational chemistry where many pieces of software do not have full-featured graphic user interfaces (GUI). Here, we present three options for accessing the command line, in the order that we recommend them.

1. Accessing the Linux command line for mac and Linux users.

- 1) Launch the "Terminal" application. That's it! Mac and Linux are Unix-based operating systems and include the same command line by default

2. Windows 10+ (WSL)

Windows introduced Windows Subsystem for Linux (WSL) in 2016, enabling users to access a Linux terminal more naturally through Windows.

It is available for Windows 10 (version 2004, Build 19041) or higher.

You may find full instructions for installation [here](#), but the gist of it is as follows:

- 1) Search for the Ubuntu app on your machine to see if it's already installed
- 2) If the Ubuntu app is not installed, download the free app from the Microsoft Store.
- 3) Launch your Linux terminal by opening the Ubuntu app. If you get errors at this step, you may need to install the Ubuntu subsystem in your command prompt (as admin) by performing `wsl --install`. You will then need to restart your machine.

The other option is to use the PuTTY app (see below).

3. Older Windows users (SSH Client)

Older versions of Windows don't feature an easy way to gain access to a Linux terminal on your local machine. You may manually install WSL, following the instructions [here](#), or you may use an SSH client to access MIT's Athena system, which is Linux based.

- 1) Download the an SSH Client such as [XShell](#) or [PuTTY](#).
- 2) Use the client to SSH into MIT's Athena system or directly into Expanse.
 - a. Open the SSH client and add the hostname (domain) `athena.dialup.mit.edu` with your Kerberos ID as the username.
 - i. For PuTTY: There should be a window with the header "Basic options for your PuTTY session". In this window, there will be a line for specifying a host name. The hostname is `<Kerberos ID>@athena.dialup.mit.edu`. The first part (`<Kerberos ID>`) is your username and the second part (`athena.dialup.mit.edu`) is the domain of the host. Double check at the "Connection" tab, "TCP" is enabled (checked). Do not adjust the port, and make sure the connection type that is selected is SSH (the default).
 - b. Open the connection, which should prompt you for DUO authentication after you enter your Kerberos password.

4. Logging into the SDSC Expanse as part of ACCESS. ACCESS is a program offering computing resources for scientific applications. In this class, we will be using their Expanse supercomputer for many of our future labs.

- 1) To login to ACCESS (you will be prompted for two factor authentication):
`ssh <access username>@login.expanse.sdsc.edu`
Hint: We will use this computer font to indicate a command or source code. When we want you to type this command at a prompt, we will color it green to help guide you. The brackets (`<>`) indicate that you will have to supply a part of the command that is unique to you, in this case your ACCESS username. When you type in your ACCESS username, do not include the brackets
- 2) To log out of Expanse/ACCESS:
`exit`

5. Getting Familiar with the Linux Commandline. A few basic commands to navigate the Linux command line are `ls` (list the files in the current directory), `cd` (change directory into

the specified directory), and **pwd** (print working directory). Basic text files can be created or edited in the command line using programs like **nano** or **vim**. Vim is more popular among experienced command line users, but we'll start with nano because it is more user-friendly. If you are an experienced command line user, use whichever text-editing command you prefer.

- 1) List the files in your current directory by typing:

```
ls
```

The output of ls should resemble what is shown in *Window 10*, but the color of the background and text may be different depending on your settings.

Hint: You can get colors with **ls -g** on most operating systems.

Window 10. The output of the terminal using the "ls" command. This prints the files contained in the home directory. In this example, we see common folders like "Documents," "Downloads," "Pictures," etc.

- 2) Now let's try changing our working directory to one of the directories shown by **ls**. Documents a common folder on most computers, so is a good example. Type:

```
cd Documents
```

Hint: When you are typing the name of a file using the command line, you can use the **tab** key to automatically complete the name of the file. This is particularly helpful to avoid typos when using long file names.

- 3) Next, let's take a look at the "path" to the current working directory. Type:

```
pwd
```

The output of pwd will be the directory you currently are in, e.g., **"/Users/timthebeaver/Documents"**. The first part of this path (**/Users/timthebeaver**) corresponds to the location of my home directory. This will look a bit different on your computer depending on the operating system. The second part of this path corresponds to the "Documents" folder, which we changed into during step 2.

- 4) If we want to change directories back to our home directory, we have three options. One way to accomplish this would be to use **".."** which the command line interprets as "up one level in the directory tree":

```
cd ..
```

or we can **cd** into a directory by specifying its exact path

```
cd <home directory path (for example, /Users/timthebeaver)>
```

or if we want to go into the last directory we were just in, we can take a step back with:

```
cd -
```

In any case, if you try the **pwd** command again, you should see that you are no longer in the Documents folder.

Hint: In the Linux command line, the **~** symbol functions as a shorthand for your home directory. To return to your home directory at any time use: `cd ~`

6. Editing plain text files and creating basic scripts with the command line.

- 1) Navigate to a file folder where you would like to save a few files and will be able to find them later. Popular choices might be "Documents" or "Desktop." For example:

```
cd Desktop
```

- 2) Now we'll create a basic plain text file using nano. Start the program by typing:

```
nano
```



Window 11. The terminal with the "nano" program open and running.

- 3) Create a basic plain text file by typing "Hello World!" and then exit nano with "ctrl+X." Nano should prompt you to ask if you want to save your changes. Type "y" to save your changes. Next, nano should prompt you to name the file. Type "hello_world.txt" and then press enter.
- 4) There should be a new file in the folder called "hello_world.txt" (visible if you use the "ls" command). We can view the contents of the file with the "cat" command

```
cat hello_world.txt
```

- 5) Now let's look at a basic bash script. Download "**example_script.sh**" from Canvas under **Modules/Lab 2: Geometry optimizations/example_scripts.sh**. Place it in your current working directory. This will be the Desktop if you used the suggested location in step 1.

Hint: If you're working on Athena or otherwise have difficulty downloading a file from Canvas, you can copy and paste the text of the script into nano and save it as a file.

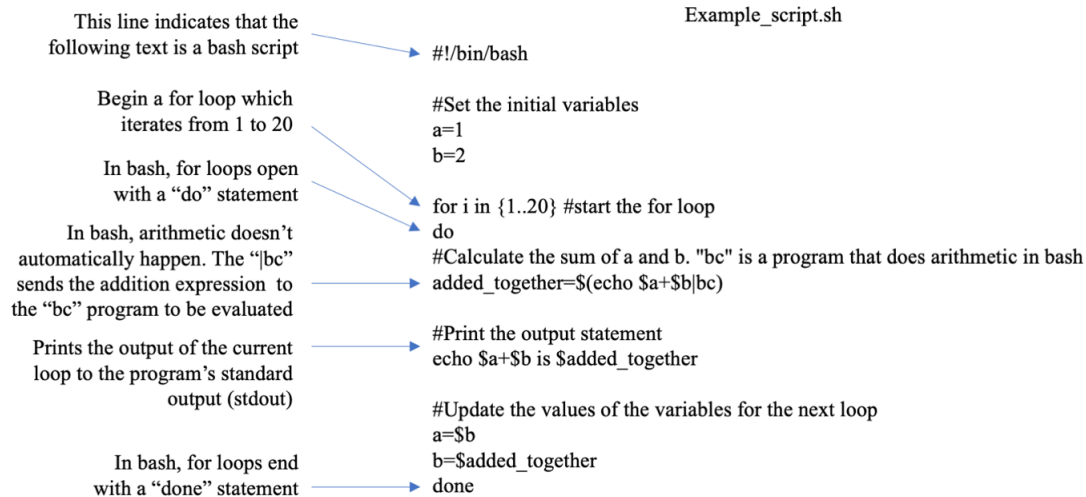


Figure 1. The contents of the example bash script, with more detailed comments.

- 6) "example_script.sh" should be in the current working directory. Run it with:

```
bash example_script.sh
```

The script should print a series of 20 arithmetic statements. In each case, the two most recent numbers are added together to generate a third number.

- 7) The output of processes at the Linux command line can be directed to a file with a ">" symbol. For example, try:

```
bash example_script.sh > example_output.txt
```

The output of the script that previously appeared in the terminal should now be contained in the `example_output.txt` file. Take a look at it with `nano`!

- 8) Use `nano` to modify the initial 2 numbers in the `example_script.sh` script, which correspond to the "a" and "b" variables near the top of the script. Instead of **1 and 2**, set these variables to **3 and 4**. Run the script again and record the output.

Worksheet prompt: What is the last number reported when these variables are **3** and **4**?