MD Workflow Single System Tutorial (LINUX OPERATION – Local Execution)

Written by Pek Ieong

The purpose of this tutorial is to introduce the Amber GPU Molecular Dynamic (MD) Kepler workflow developed by NBCR (http://nbcr.ucsd.edu/wordpress2/) and WorDS (http://words.sdsc.edu/) through the support from NVidia (http://www.nvidia.com/). This workflow seeks to automate the Amber MD job submission in either a GPU cluster or on a local machine. It serves as a tool to run Amber Molecular Dynamic for experienced users and an introduction to Amber and Kepler for new users. The workflow is built using the Kepler platform. To start, a local installation of the latest Kepler software is required.

DISCLAIMER
To run this workflow, users need the AMBER Molecular Dynamic Package (http://ambermd.org/) with a workable MPI pre-installed in the user's local computer. For more information about the MPI that can work with AMBER software package, please consult the Amber website.

Users who are not familiar with Kepler, please start from Step 1.
Users who have experience with Kepler and have Biokepler 1.1 and Kepler – 2.4.2 installed in user's computers, feel free to start from Step 4

1. Installing Kepler:
   Users can download the latest Kepler software from the following link:
   https://kepler-project.org/users/downloads

   To install on **Windows**, double click on the executable Kepler-2.4-win.exe.
   To install on **Mac OS X**, mount the Kepler-2.4.dmg by double-clicking on it and drag-and-drop Kepler-2.4 folder into the /Applications folder.
   And to install on **Linux**, simply untar kepler-2.4-linux.tar.gz at a location of your choice, which will generate a folder named kepler-2.4. The command line to untar a tar.gz file:

   tar -C /myfolder -zxvf kepler-2.4-linux.tar.gz

   For the following tutorial, we will work on a linux platform as an example.

2. Starting Kepler:
   Move into the Kepler installation directory

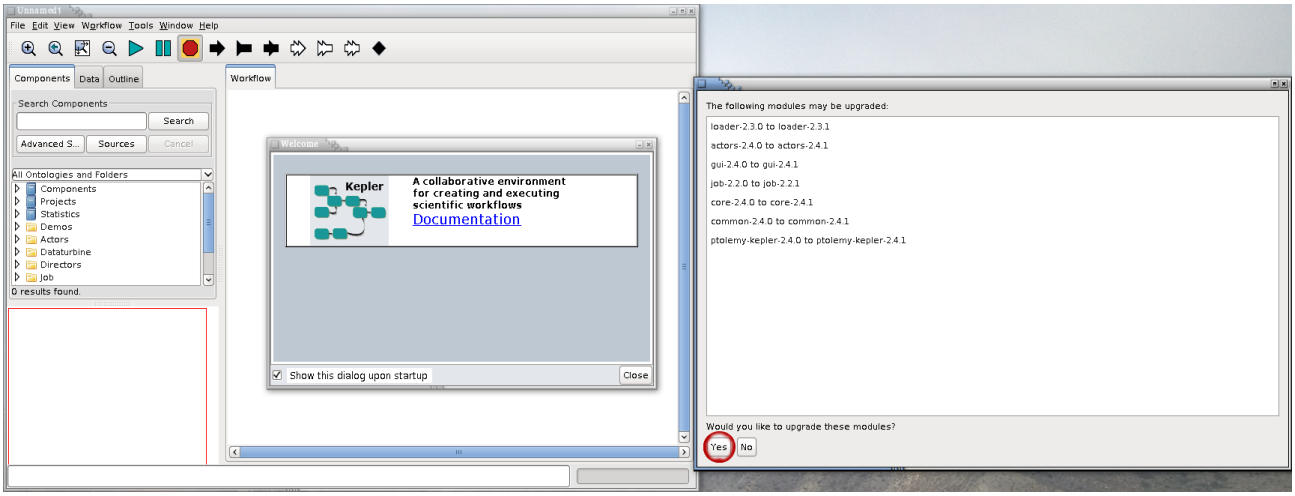
   cd /myfolder/kepler-2.4

   To start the Kepler GUI, execute the kepler.sh script.

./kepler.sh

A pop up window will prompt first time users to upgrade several modules (red circle in Fig 1) and Kepler will automatically upgrade and restart.

Fig 1.

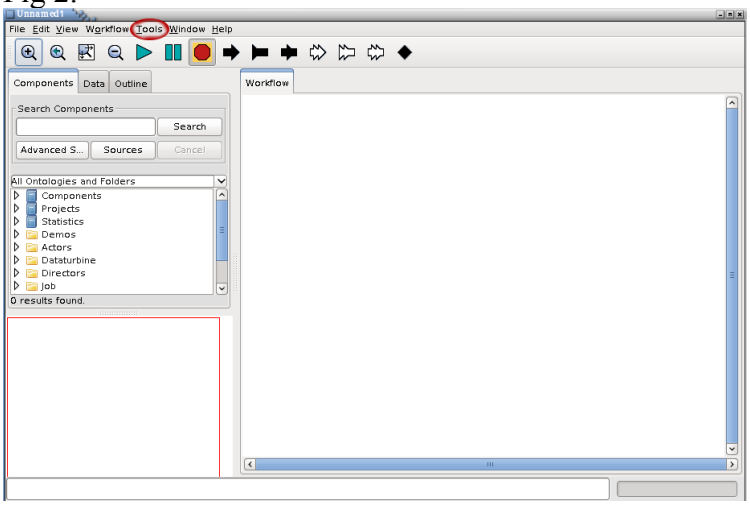


3. Installing Biokepler:

Our workflow allows users to choose between running the MD simulations with either a GPU cluster or a local GPU card. To allow this flexibility in a single workflow, we utilized a module from Biokepler. Hence, besides installing the Kepler software, we need to also install the Biokepler module.

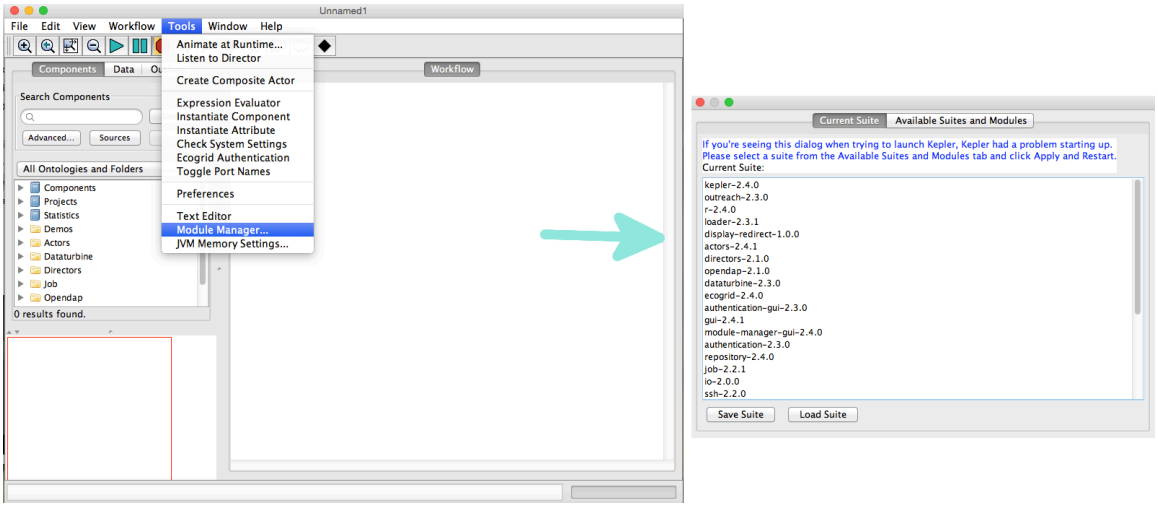
The Kepler GUI will reopen after the modules are successfully updated (Fig 2). Go to the menu bar on the top left of the screen and select Tools (Fig 2, red circle).

Fig 2.



From the drop down menu under Tools, select Module Manager and a new window will pop-up (Fig 3).

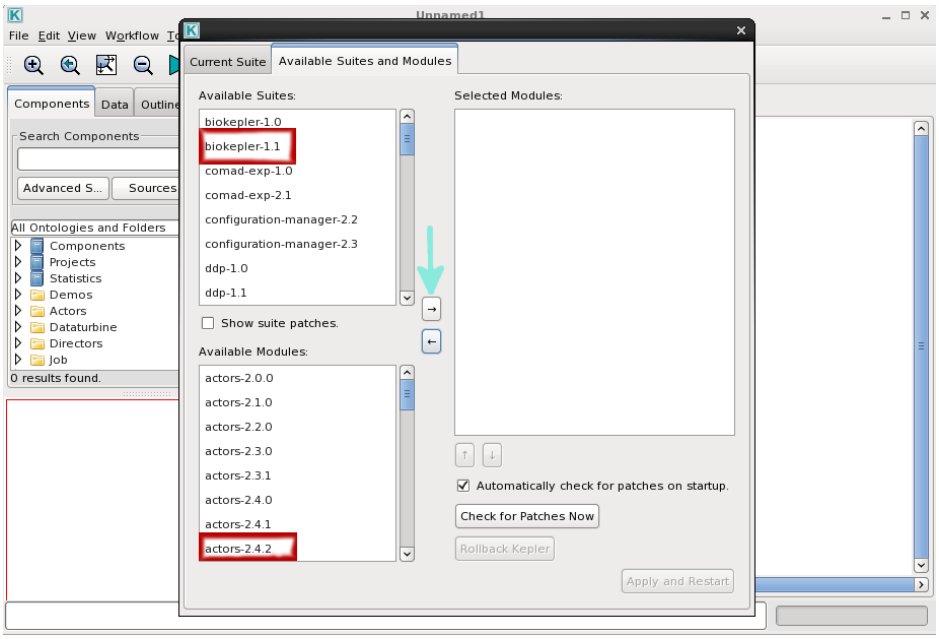
Fig 3.



There are two tabs on the top of the new window, click on Available Suites and Modules, which will show a list of available suites and modules. Among the Available Suites, select biokepler-1.1, and then click on the right arrow to transfer biokepler 1.1 to the Selected Modules window (Fig 4).

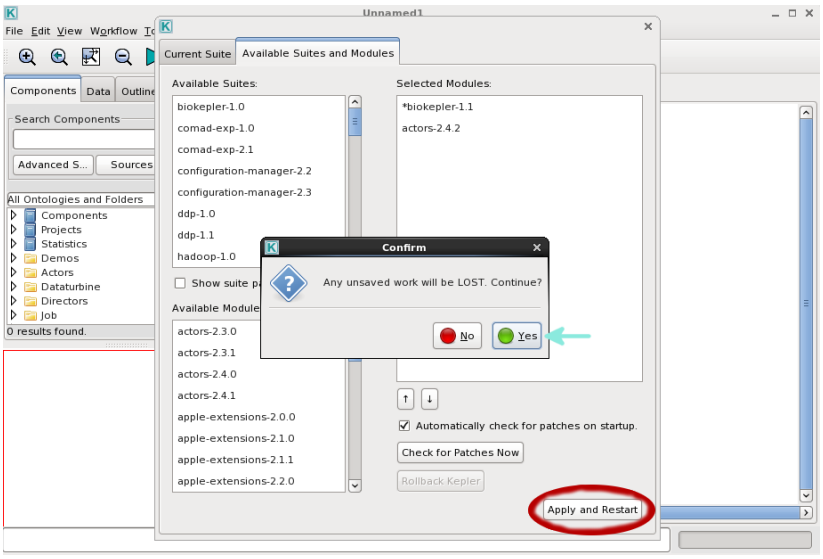
Do the same thing for the actors-2.4.2 modules. Among the Available Modules, select actors-2.4.2, and then click on the right arrow to transfer actors-2.4.2 to the Selected Modules window (Fig 4).

Fig 4.



Finally, apply the selected modules, and restart Kepler by clicking on the Apply and Restart tab at the bottom right corner of the window (Fig 5). Kepler will warn users that any unsaved work will be disregarded. Please save any unsaved work, and click Yes (Fig 5).

Fig 5.



The Kepler GUI will reopen after successfully installing Biokepler. You can close it because we will run MD workflow using command prompt.

4. Downloading the MD workflow:

   The MD workflow is prepared to run on the Kepler platform. The workflow itself is a product developed by NBCR (http://nbcr.ucsd.edu/wordpress2/) and WorDS (http://words.sdsc.edu/). The MD workflow, named MD_WF_SingleSys.xml, can be downloaded the Amarolab website under resources and inside Amber GPU Molecular Dynamics (MD) Workflow Tool and Tutorial or from the following link: (https://amarolab.ucsd.edu/mdworkflow/mdworkflow.html).

5. Setting up input folders:

   Before running the MD workflow, you need to set up and organize input folders in a specific required format. To run the MD workflow, users will need topology and coordinate files for the system of interest as well as Amber input scripts for minimizations and molecular dynamics (MD) simulations. We generated a sample folder with required files for a test system downloadable from the Amarolab website under resources and inside Amber GPU Molecular Dynamics (MD) Workflow Tool and Tutorial or from the following link: (https://amarolab.ucsd.edu/mdworkflow/mdworkflow.html).
   Inside this link, there is an Input_File.zip file. Users can find both the topology and coordinate files and all of the MD input scripts inside (Table 1).

Table 1.

Inside Input_Files:

p53_wt	p53_wt.top, p53_wt.crd
confDir	min1_switch.conf, min2_switch.conf, min3_switch.conf, min4_switch.conf, min5_switch.conf, md1_switch.conf, md2_switch.conf, md3_switch.conf, md4_switch.conf, md5_switch.conf

Download the files and create a new directory.

mkdir MD_TEST

Then, move the zip file into the new directory.

mv ~/Downloads/Input_Files.zip

Unzip the Input_Files zip file and then remove it from the directory.

unzip Input_Files.zip
rm Input_Files.zip

Two folders named confDir and p53_wt will now appear in the empty directory. In order to run the workflow, we need a folder containing all the Amber input files including five steps of minimization (min1 to min5), one step of heating (md1), three steps of equilibration (md2 to md4), one 1ns of productions (md5) using the MD workflow. In the other folder named p53_wt, it is where users will put the topology and coordinate files. The folder name needs to match the stem name of the topology and coordinate file. For example, if the two files are named apple_tree.top and apple_tree.crd, the folder name will need to be apple_tree.

Now, you are ready to run the MD workflow for a single system.

6. To run MD Workflow.

   In your local computer, go to the directory where Kepler was installed and where kepler.sh is located.

   cd /PATH/kepler-2.4

   And type the following command line in that directory to start the MD workflow. Users will need to modify the command based on their needs. The definition for each parameter is listed below. Users are REQUIRED TO type in the absolute path for each file when running Kepler. If an absolute path is not provided when inputting the command in command prompt, Kepler won't be able to find the files or folders of interest.

   The general command to run Kepler through command prompt is

   ./kepler.sh -runwf -nogui -force -parameters values /PATH/workflow.xml

Here is an example of how Kepler can be run for our test case.

./kepler.sh -runwf -nogui -force -ExecutionChoice LocalExecution -NP 8 -inputFolder [see parameter] -CMPD p53_wt -AMBERHOME_local $AMBERHOME /LOCATION/OF/THE/MD/WORKFLOW/MD_WF_SingleSys.xml

Parameters:

Please customize below parameters for your execution:

ExecutionChoice: This parameter gives option to run same workflow on a local machine or a GPU Cluster. For local execution: LocalExecution
NP: number of processors for minimization step. For example: 8
CMPD: The folder name containing the topology and coordinate files. For example: p53_wt
AMBERHOME_local: $AMBERHOME on your local machine. For example:
inputFolder: The inputFolder path on your local machine.

At the end, when the jobs are successfully submitted, users will see similar outputs in their terminal (Fig 6.) Kepler will keep track of the job status and close when the jobs are finished.

Fig 6.