User documentation for DROIDS 3.0+maxDemon 1.0 – a machine learning assisted GUI-based pipeline for comparative protein dynamics

Gregory A. Babbitt T.H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester NY USA

author email address: gabsbi@rit.edu

please report bugs to this email

System Requirements – Debian Linux desktop OS with 1 or more GPUs. Linux Mint 18/19 is recommended with Nvidia GTX 1080 or larger. Be sure to also check the Linux Mint 'Driver Manager' after initial build and install all recommended Nvidia drivers.

NOTE – software can be most easily installed by running 'perl DROIDS+AMBERinstaller.pl' a perl installer script included with our download folder or GitHub repo. After installation, the software runs as 'python DROIDS.py' on the Linux terminal opened from within the DROIDS folder

Software – Amber16/18, AmberTools 16/18, UCSF Chimera 1.11 or 1.13 (additionally ChimeraX optional), CUDA 8.0/9.0, CUDA toolkit, perl-tk, python-tk, and R. (Note: Amber18 install on Linux Mint 19 will likely require setting up older versions of the gcc, g++, and gfortran compilers. Version 5.0 works well. Our installer will lead the user through this process if needed. Do not use CUDA 7.0 or earlier nor version 10.0 or later.

Debian and python packages – gedit, grace, gdebi, gparted, evince, perl-tk, python-tk, python-gi, gstreamer (and dependencies), and these Amber dependencies (csh flex patch gfortran g++ make xorg-dev bison libbz2-dev). If you plan to use VR, install Steam, SteamVR and vulkan library.

Perl packages – Statistics::Descriptive

R packages – ggplot2, gridExtra, dplyr, caret, FNN, e1071, kernlab, class, MASS, ada, randomForest, CCA, CCP, parallel, rpsychi, foreach and doParallel.

NOTE: We supply an installer script (DROIDSinstaller.pl) with download. It can setup the whole system (including CUDA, Amber18, UCSF Chimera, and R) on a fresh Linux Mint install or Virtual Machine. It can also skip over these if already installed (i.e. only will install package dependencies).

DROIDS download - https://github.com/gbabbitt/DROIDS-3.0-comparative-protein-dynamics
Get the most recent release and download as .tar.gz file and decompress

INSTALLATION

1. Setup a fresh Linux Mint OS and check the Driver Manager to properly install Nvidia drivers.
2. Untar and Open DROIDS folder and copy DROIDSinstaller.pl to your desktop
3. Open terminal at the desktop and run installer (perl DROIDSinstaller.pl)
4. Follow directions. NOTE: the installer will install AMBER, UCSF Chimera, and R as well as all the system dependencies and packages required by DROIDS+maxDemon. If, AMBER, R and Chimera are already installed, you can skip these segments of the installation. Also note that AMBER 18 installation requires older gcc, g++ and gfortran compilers than are downloaded with Linux Mint 19, so we include instructions at the terminal for how to reset these to older versions for proper AMBER install.

To Run DROIDS

Copy the downloaded DROIDS folder and rename it more simply, with reference to your analysis (e.g. DROIDS_1ubq_temperature). You will also later add the .pdb files you want to analyze into this folder.

To run DROIDS, open terminal from within the folder and type python DROIDS.py

or type perl DROIDS.pl (for older versions v1.0 and v2.0)

Upon, first run on a given computer, you will need to set the paths to various software on your system (i.e. Amber and Chimera) into a text file named paths.ctl. After you pass the GPL license information, a GUI will open asking you for these paths. The most likely paths will also be flashed to your terminal after running automatic searches. Use these paths on first run and double check them. If these softwares do not run later when called by DROIDS, it is likely that these paths are not fully specified. Once you have successfully created a paths.ctl file for your computer, save it somewhere safe and copy it into any new DROIDS analyses folders you set up. Then you can skip this step in future runs.

INSTRUCTIONS FOR DROIDS v3.0 ON WINDOWS and/or GOOGLE CLOUD PLATFORM

Option 1: running a Linux Mint Virtual Machine (VM) on Windows PC

1. Download and install VirtualBox from this website

   https://www.virtualbox.org/

2. Get a .vdi file (for Linux Mint) from this website

   https://www.osboxes.org/virtualbox-images/

3. Build a Linux Mint VM following instructions from VirtualBox
4. Copy the required files to the VM (i.e. DROIDS-3.0.tar.gz, Amber18.tar.gz, AmberTools18/19.tar.gz, Chimera-1.14-linux_x86_64.bin). These can be obtained from the following websites
   https://people.rit.edu/gabsbi/
   https://ambermd.org/
   https://www.cgl.ucsf.edu/chimera/download.html
   and you'll probably want Modeller enabled on Chimera

https://salilab.org/modeller/

5. Run the DROIDS+AMBERinstaller.pl script on the VM's desktop terminal

   perl DROIDS+AMBER_installer.pl

   NOTE: your PC must already have Nvidia GPU hardware, CUDA and Nvidia graphics drivers properly installed.

Option 2: running an Ubuntu Linux VM on GCP (Google Cloud)

https://cloud.google.com/

1. Open GCP account, go to your console and request resource quota to enable building GPU VM instances. (e.g. use 8CPUs and add 1 V100 GPU)
2. Upload the required files to a Google Cloud Storage Bucket (i.e. DROIDS-3.0.tar.gz, Amber18.tar.gz, AmberTools18/19.tar.gz, cuda-repo-ubuntu1704-9-0-local_9.0.176.deb (or an equivalent version of cuda), Chimera-1.14-linux_x86_64.bin). These can be obtained from the following websites
   https://people.rit.edu/gabsbi/
   https://ambermd.org/
   https://developer.nvidia.com/cuda-downloads
   https://www.cgl.ucsf.edu/chimera/download.html
   and you'll probably want Modeller enabled on Chimera
   https://salilab.org/modeller/
   make sure the permissions on all files are executable (e.g. sudo chmod +x 'filename')

3. On your Dashboard, go to 'Compute Engine' and build your VM instance. IMPORTANT: Be sure to add an Nvidia GPU and use Ubuntu 16.04LTS or 18.04 LTS so that you can install and link from a remote desktop application.
4. Once the VM appears, connect to it via the SSH link and run the following commands.
   sudo passwd (to reset root password)
   sudo passwd yourusername (to reset your user password)
   sudo apt-get install xrdp
   sudo apt-get install xfce4
   sudo service xrdp restart

5. Leave this terminal open and go to your Windows RDP (Remote Desktop) application and enter the external IP address from your VM to open the new xfce desktop you installed with your new passwords.
6. At this point you can transfer your files from your cloud bucket or your own computer to the VM home folder (using the 'gear' icon on the ssh terminal of your VM). Open the home folder and move your files to your desktop and then proceed with the installation using the DROIDS+AMBERinstaller.pl script at either the desktop or SSH terminal.

   perl DROIDS+AMBERinstaller.pl

NOTE: when the script pauses and asks for information typed into secondary terminals and the .bashrc file, these files and terminals may need to be opened manually at the VM instance SSH link or else on the remote desktop. If R package installations fail after installing R, open R manually at the command line (i.e. type 'R') and then paste or type the following

>install.packages(c('ggplot2', 'gridExtra', 'dplyr', 'caret', 'FNN', 'e1071', 'kernlab', 'class', 'MASS', 'ada', 'randomForest', 'CCA', 'CCP', 'doParallel', 'foreach', 'rpsychi'))

then to exit R

>q()

Improvements and upgrades over previous versions

To enhance the user experience and scientific utility, DROIDS v3.0 offers many new features beyond earlier major release versions 1.2 and 2.0. These are summarized below.

- New GUI organization directs users to specific comparative tasks/applications in Table 1
- A new control file builder for managing path dependencies in Linux is included
- Amber16/18 support has been beta tested and is defined via paths.ctl file
- Single or dual GPU user options are available for faster analyses
- Automated structure prep (dry and reduce) via pdb4amber is now included in the GUI. The 'reduce' variable is optional allowing users to either setup their own protonation states ahead of DROIDS, or simply allow DROIDS to hydrogenate the input structures entirely.
- Program/package dependency installer script named 'DROIDSinstaller.pl' is included. It will lead users through all dependencies required after a fresh Linux build, including CUDA libraries and tools required for Nvidia GPU accelerated Amber in the Linux environment
- KL divergence (= relative entropy) definition of dFLUX is now included as an option providing a richer color mapping of dFLUX in images and movies than the simple averaging algorithm offered in earlier DROIDS versions
- Binding interaction analysis for both protein-DNA and protein-ligand systems is now offered with dedicated GUI for these comparisons. Protein-ligand system setup includes QMMM preprocessing in Antechamber and SQM.
- LeAP control files for explicit solvent runs are now presented for advanced user modifications (e.g. changing ion concentration, water model, water box dimension of volume).
- Dedicated GUI allowing genetic mutation placement (on DNA or AA) are included for setting up variants to analyze
- Self-stability and temperature shift analysis has its own dedicated GUI, allowing users to copy the input pdb file to compare MD ensembles generated on identical structures at the same of at different temperatures
- MaxDemon 1.0 - machine learning based detection of functionally conserved dynamic regions
- MaxDemon 1.0 - machine learning based impact assessment of variants (genetic, structural or binding)
- Dynamic visualization and movie rendering of machine learning classification performance
- Virtual reality and ChimeraX compatibility is also supported (additional information and download code can be found here https://cxtoolshed.rbvi.ucsf.edu/apps/moleculardynamicsviewer https://github.com/kdiller713/ChimeraX_MolecularDynamicViewer

Current bugs –

1. In some systems, when the maxDemon machine learning deployment button is used for the first time, an error 'can't open temp test file' is thrown. If the user closes the maxDemon GUI and reopens it at the Linux terminal ('perl GUI_ML_DROIDS.pl'), the button will work as intended and learners can then be deployed.

NOTE: If you use DROIDS for published work please use the following citations

Babbitt G.A. Fokoue E. *Evans J.R. Diller K.I. Adams L.E.* 2020. DROIDS 3.0 - Detection of genetic and drug class variant impact on conserved protein binding dynamics. BIOPHYSICAL JOURNAL 118: 541-551 CELL Press.

Babbitt G.A. *Coppola E.E. Mortensen J.S. Adams L.E. Liao J. K.* 2018. DROIDS 1.2 – a GUI-based pipeline for GPU-accelerated comparative protein dynamics. BIOPHYSICAL JOURNAL 114: 1009-1017. CELL Press.

DROIDS was produced by student effort at the Rochester Institute of Technology under the direction of Dr. Gregory A. Babbitt as a collaborative project between the Gosnell School of Life Sciences and the Biomedical Engineering Dept. Visit our lab website (https://people.rit.edu/gabsbi/) and download DROIDS from Github at https://github.com/gbabbitt/DROIDS-2.0---free-software-for-comparative-protein-dynamics

Implementation

It is strongly advised that users be comfortable with how to prepare PDB files for molecular dynamic (MD) simulation using GPU accelerated AMBER 16/18 (pmemd.cuda). DROIDS assists with modifying .pdb files named in the GUI for AMBER simulation, however the user should become very familiar with the programs running at these steps (i.e. antechamber, pdb4amber, and teLeap) and read through all output at the DROIDS terminal to ensure that the structures are properly prepared for MD simulation. You must consult the AMBER documentation for this knowledge. The DROIDS GUI provides automation of teLeap, a program for pdb file setup, but care must be taken to read output on the Linux terminal for any errors. The programs 'antechamber' and 'pbd4amber' are used by DROIDS in modifying files for MD and are generally prior to starting teLeap in DROIDS. Please consult the Amber16 user manual for more details. Typically preparation includes (A) removing mirrored images and other chemical artifacts (done manually in Chimera prior to DROIDS), (B) performing a structural alignment (using Chimera MatchMaker and Match->Align when prompted by DROIDS) followed by subsequent saving of a Clustal format file (.aln), (C) adding H atoms and removing crystallographic waters (use pdb4amber button in DROIDS to dry and reduce), (D) estimating and loading force field parameterization regarding important ligands if a protein-ligand interaction is modeled (use antechamber button). Then finally (E) run teLeap button in DROIDS to

setup topology and coordinate files for simulation. For v2.0 we have added script to check the file sizes of teLeap output files and recommend whether the process likely failed of succeeded at this step. teLeap is nicely verbose, so warnings on terminal when running teLeap button is very helpful for any indications of problems specific to your structural models. For many at this stage of model prep, it is not unusual to go back to modify the original .pdb file and run through the prep stages again. Be sure to view your models in Chimera using the 'all atom' preset so that you do not miss small molecules that might trip up the MD setup. Amber is designed not to run unless all atoms in your system can be properly parametrized by the force field you have chosen. Many force fields are available to try in the amber16/dat/leap/cmd folder. Many are appropriate only for certain macromolecules, and analysis of binding interaction will require several are loaded. ALSO NOTE: AMBER 16/18 software must be licensed from the University of California. More details about purchasing and installation can be found at http://ambermd.org/. DROIDS is tested on Linux Mint 18.1 and Ubuntu 16.04 and is offered freely under the GPL 3.0 license and is available on GitHub https://github.com/gbabbitt/DROIDS-1.0

DROIDS is activated by entering 'python DROIDS.py' at the Linux terminal opened from within the DROIDS folder. DROIDS v3.0 initially starts with a small GUI requesting user to add paths to Chimera and Amber's force field data files (e.g. amber16/dat/leap/cmd). As Amber16/18 is typically installed to the Desktop, this path will be different on different machines. Make sure you edit the path appropriately before attempting to run DROIDS. The GUI will create a paths.ctl file. Once this file is created for your individual machine, it can be saved and dropped into DROIDS folders prior to each run. The typical bashrc file can be used similarly, but this GUI was added to make this initial setup simpler for less experience Linux users. Once the paths GUI is closed, the main DROIDS v3.0 GUI will appear. Here the user is directed to choose one of the various types of comparative analysis that can be done, choose MD sim software, and indicate whether the machine is running a single or dual GPU. Upon clicking 'run DROIDS' the user is taken to the first main GUI for setup, running MD, and parsing of MD simulation output. The second main GUI controls the DROIDS statistical analyses and the last main GUI controls the image color-mapping and movie rendering and viewing options.

SEE OUR TUTORIAL (PDF) FOR MORE EXPLICIT INSTRUCTIONS ON RUNNING DROIDS

IMPORTANT NOTE: When running DROIDS on many protein comparisons, we find that explicitly solvated systems (i.e. PME method) tend to yield better and more conservative results regarding the significance of the KS test when compared to implicitly solvated comparisons (i.e. GB method). This is likely expected due to the many more degrees of freedom under the PME option as well as its better approximation to reality. A three point solvent model (tip3p) is default method in DROIDS. This is for sake of efficiency. If a more accurate solvent is needed we recommend the users edit the .bat files that pop open when running teLeap from the DROIDS GUI. The user can manually change the references to tip3p to the tip4p, tip5p or tip6p models. Another default state of our software is to charge neutralize the

protein. The .bat files can also be edited by experienced users to alter the ion concentrations in the simulation. For more complicated setups, the numbers of ions needed for given box size and salt concentration can be determined using the method and tool cited below.

SLTCAP: A simple method for calculating the number of ions needed for MD simulation

Jeremy D. Schmit*,†, **Nilusha L. Kariyawasam**‡, **Vince Needham**†, and **Paul E. Smith**‡
†Department of Physics, Kansas State University, Manhattan, KS 66506, USA
‡Department of Chemistry, Kansas State University, Manhattan, KS 66506,

J Chem Theory Comput. 2018 April 10; 14(4): 1823–1827. doi:10.1021/acs.jctc.7b01254.

We recommend that users explore many methods of solvation when using DROIDS. Implicitly solvated protein comparisons run relatively fast and may be useful for an initial investigation of a large system, however comparison of explicitly solvated systems may yield more realistic local variation in mutational impacts.

IMPORTANT: Given that MD simulations are well known to exhibit complex and often chaotic behavior, we also strongly recommend that users of DROIDS repeat analyses of given systems in order to determine best parameter settings for ensemble size, lengths of production runs and overall reproducibility of the final results.