DROIDS 3.0 +maxDemon TUTORIAL

System requirements – Linux OS with 1 or 2 dedicated Nvidia GPUs

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, 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.

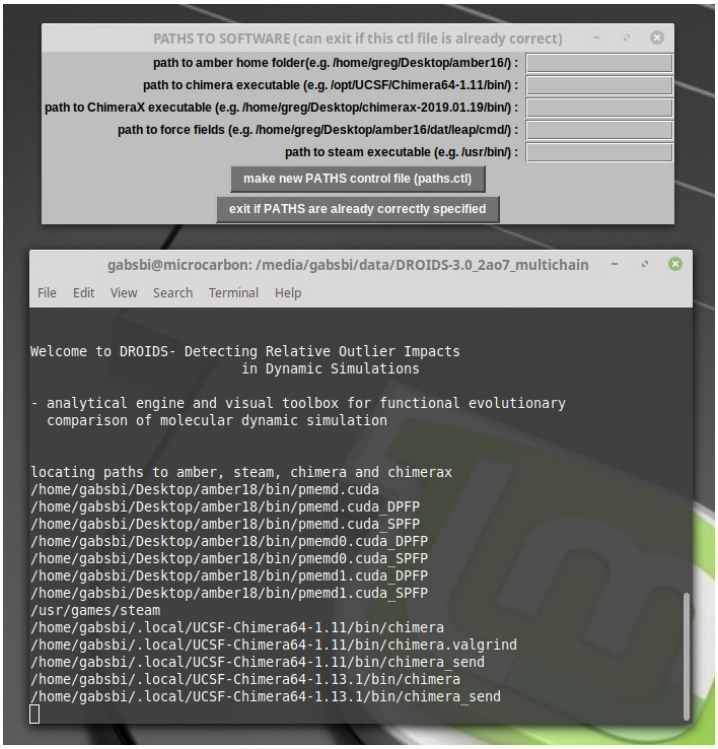
To Run DROIDS

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.

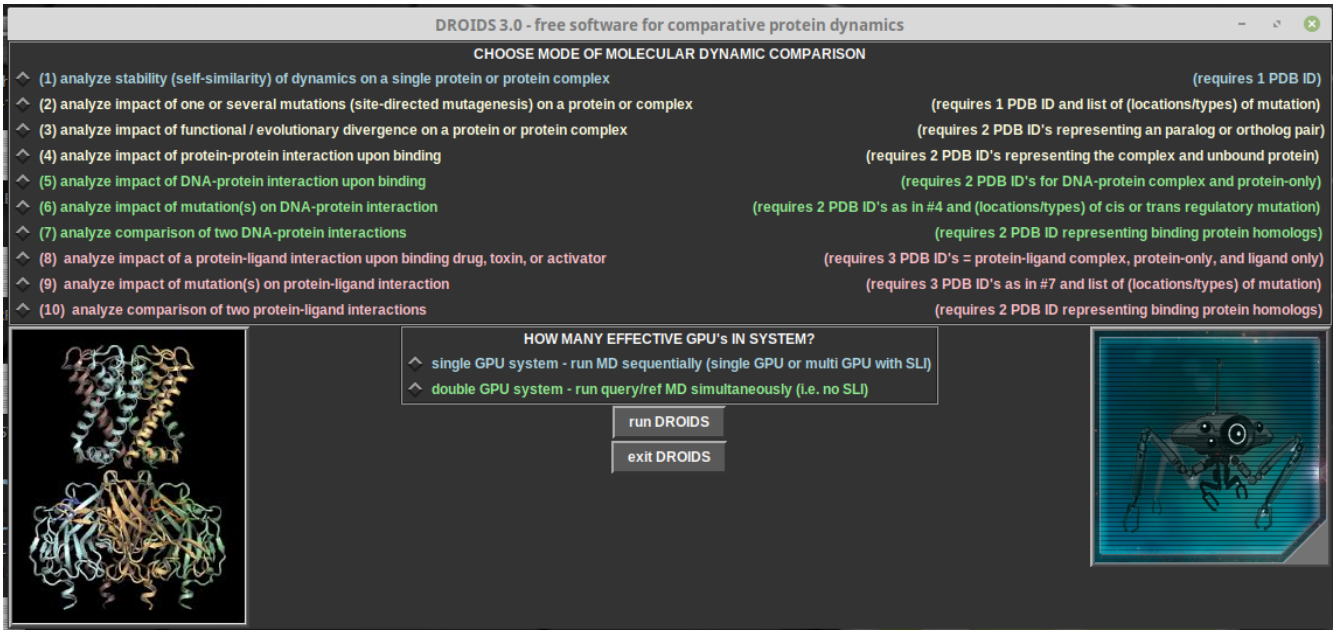
To run DROIDS, open terminal from within the folder and type 'perl DROIDS.pl'

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.

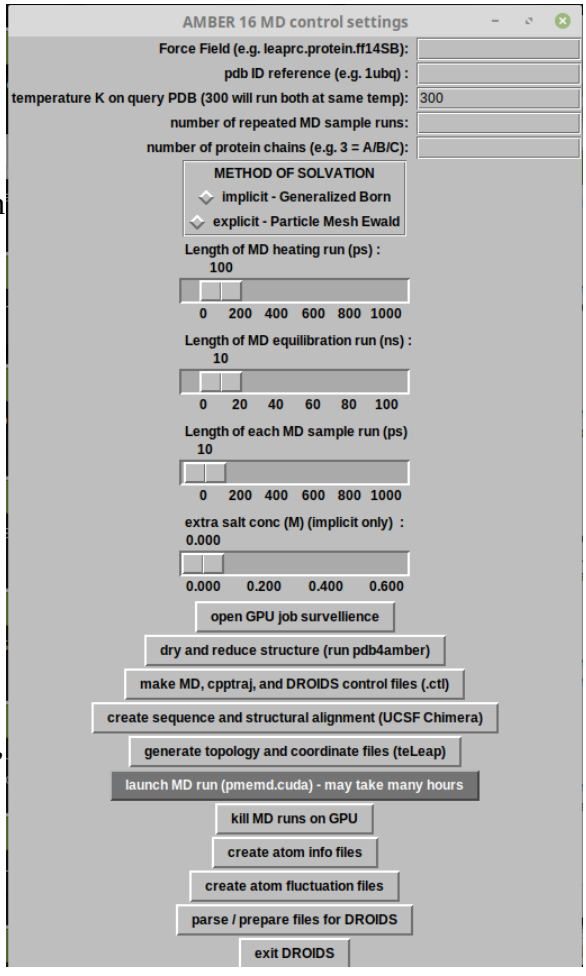
The paths for ChimeraX and steam can be ignored unless you are trying to use a VR headset. The force field libraries are very necessary and are typically in the amber18 folder on your desktop (e.g. amber18/dat/leap/cmd/)



After passing the paths GUI, you should see the GUI for the main DROIDS menu.



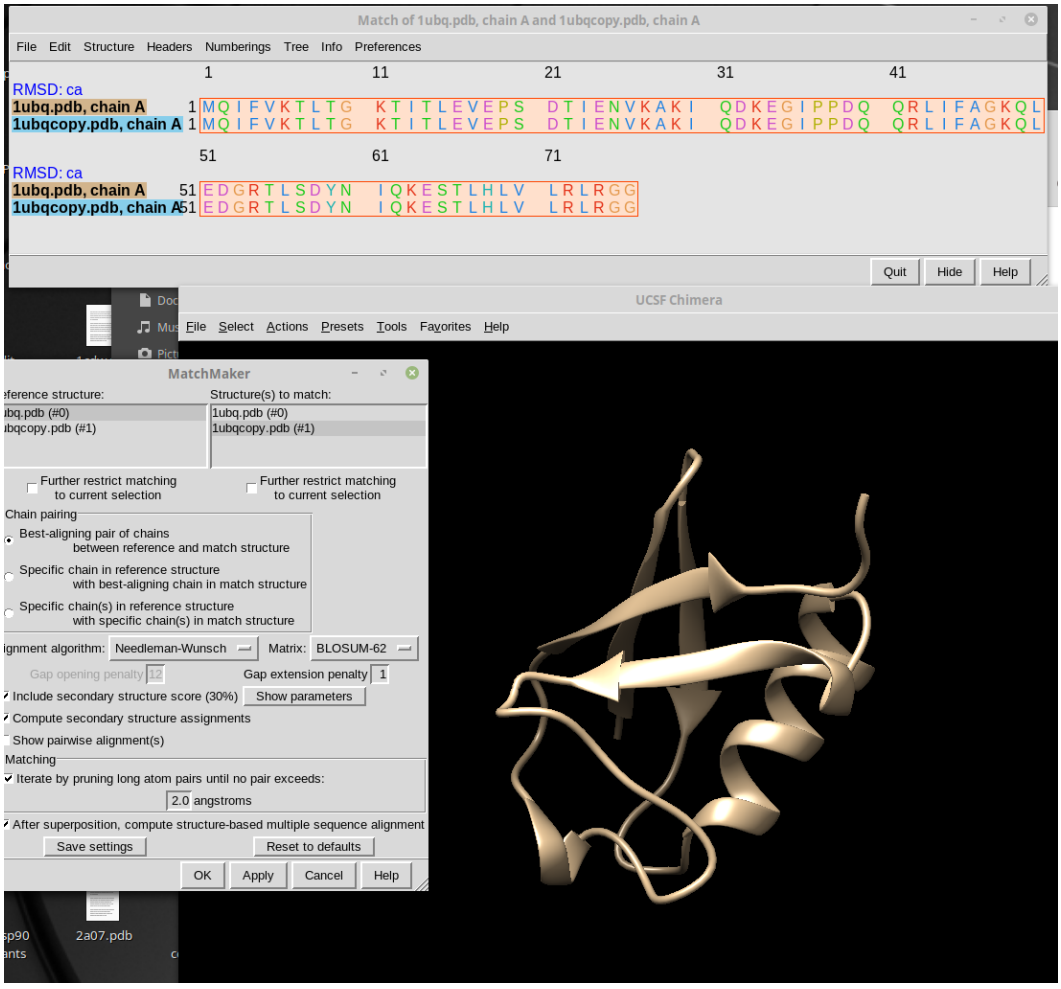
Here you will select the type of molecular dynamic comparison you want to study, and then enter the number of system independent GPU's you will use to run simulations. If you have multiple GPU's connected by SLI, then select single GPU. If you have two GPU's that are unconnected, select douple GPU system. This will run the simulations on the reference and query .pdb files simultaneously. Click run DROIDS to open the Amber control GUI appropriate to your comparison.



DROIDS TUTORIAL 1 – COMPARING UBIQUITIN DYNAMICS AT TWO TEMPERATURES

First, obtain the .pdb file for ubiquitin (PDB ID: 1ubq) from the Protein Data Bank using either the RSCB PDB website or the PDB fetch tool in UCSF Chimera. Place the file in the DROIDS folder and run DROIDS. On the main menu select (1) analyze stability of dynamics in single protein. The MD control GUI will open (image right). Here you will enter the PDB ID, protein forcefield, number of replicate MD runs (i.e. ensemble size) and other information required for the Amber MD simulation.

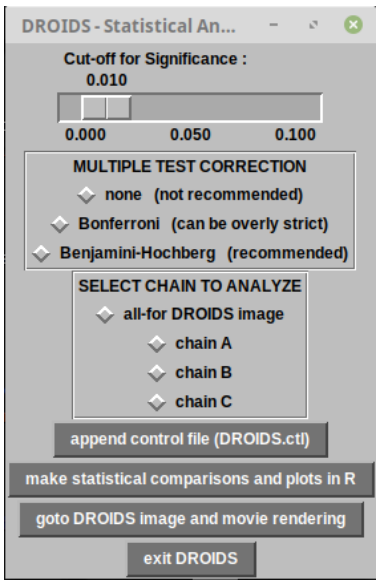
This pipeline will automatically copy your pdb file to create two identical files to analyze as a reference and query. The default temperature is 300K. The reference will always run at this temperature. The query temperature can be changed to analyze temperature sensitivity of the protein (or left the same if user wants to study self similarity of dynamics on the protein). Here, set the temperature to 250K. Use the force field 'leaprc.protein.ff14SB' for this tutorial. The 'dry and reduce' button will run pdb4amber to create reduced/hydrogenated versions of your pdb files that are absent of crystallographic waters. Answer Y for yes at terminal to run this after clicking the button. Prior to running DROIDS, the pdb file should be checked over in Chimera using the all atom preset to see if any unusual ions or small molecules are present in the crystal. They can be removed using commands like 'delete selected' or 'delete Mg' on the Chimera command line before the clean pdb file is resaved. NOTE: Amber prep will add ions back when the structure is charge neutralized prior to simulation. The settings on the GUI are stored as MDq.ctl and MDr.ctl and DROIDS.ctl when control files are created. The heating stage should be longer on larger proteins. Typical equilibration for most small proteins takes anywhere from 1 to 10 ns depending on inherent stability. Ubitquitin is very stable, so 1ns is fine. Most of our published analysis involve several hundred MD production runs at 0.5ns (500ps). Select explicit solvent. The size of solvated system and size of GPU will determine upper limit of what is possible with DROIDS on your machine. All DROIDS analyses require an alignment file for the query and reference PDB structures. This can be obtained by running the 'create sequence and structural alignment' button on the GUI. A Chimera window will open both files (1ubq.pdb and 1ubqcopy.pdb) and the user will use the MatchMaker Match-Align tools under Structural Comparison (image below) to create a good sequence alignment and save it in Clustal format (e.g. my_align.aln). Upon exiting Chimera, the terminal will prompt the user to type the name of the saved file, and DROIDS pipeline will copy it into a name convention required by future steps of analysis. NOTE: If files have multiple chains, a Clustal file for each chain will be required.

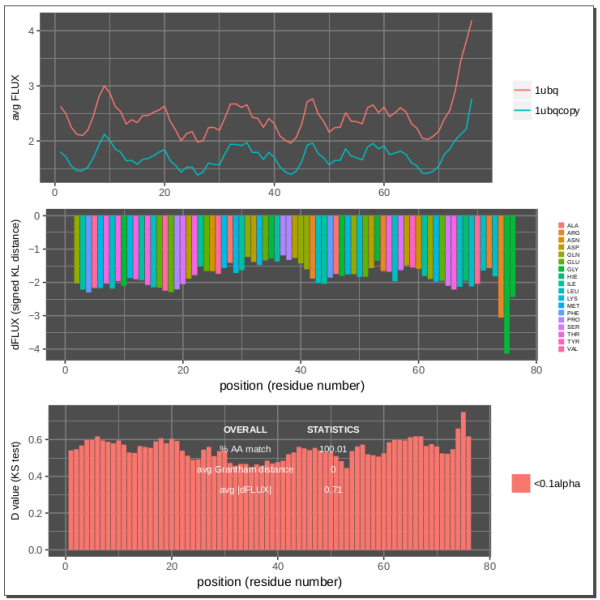


The last preparation prior to launching MD will be to run the ‘generate topology and coordinate files’ button. This runs a program in Amber called teLeAP to generate .prmtop and .incrd files that specify the initial starting conditions of the simulation. The is a complex piece of software that will check your structure against the chosen protein forcefield for any problems, will charge neutralize the system with Na+ and Cl- ions. If any atoms in your system are not recognized by the software, this step will fail. When using Amber MD on novel structures it takes practice and education about the method of MD to do this properly. Any warnings will be reported to the screen, and should be considered. During running this subroutine, a .bat text file with commands that control teLeAP will open allowing experienced users to modify the basic prep. The size of the water box can be expanded for large structures, and the default water model TiP3p can be changed. Additional ions can be added etc. DROIDS makes some basic assumptions about intended use, and so in most cases this file does not need modified and can simply be closed. At the end of the teLeAP subroutine, the file sizes of the expected resulting files are checked to see if they have been properly created. This information is reported as messages ‘teLeAP appears to have run’, or ‘teLeAP appears to have failed’ on the Linux terminal.

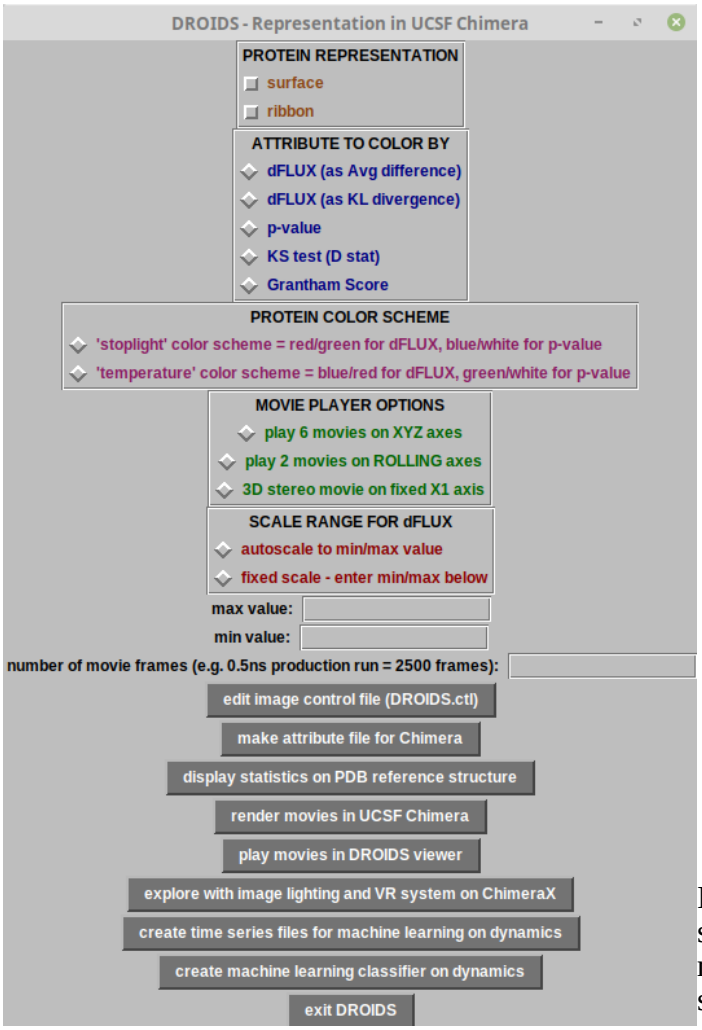
At this point, the ‘launch MD’ button can be tried. If all is well, the basic steps of MD (energy minimization, heating, equilibration and production runs) will open and run. The GPU surveillance terminals, opened earlier should show a single process pmemd.cuda running (single GPU system) or two processes pmemd0.cuda and pmemd1.cuda (double GPU system) running. The nvidia-smi terminal should show the GPU card(s) warming up and running at 100%. If user needs to stop these for any reason, they should terminate the DROIDS subroutine by type ctrl+C until ‘MD simulation complete’ appears. The user should then run pkill pmemd button to terminate processes on the GPU’s. **Failure to launch properly usually indicates that prior file preparations had failed, more rarely, GPU Linux drivers are not working properly on the system (try a reboot), or in very rare instance, GPU hardware has failed.** Progress on MD simulations can be monitored at the Linux terminal(s).

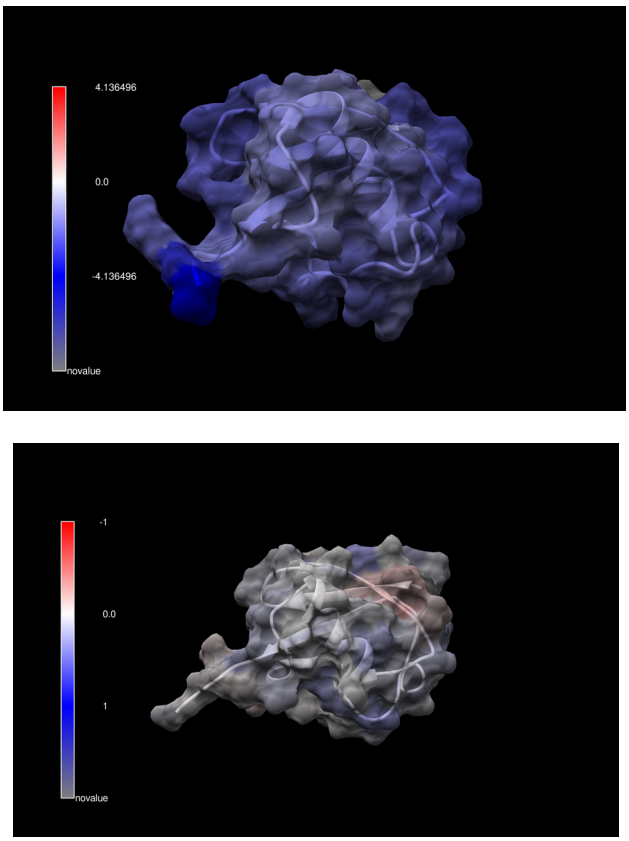
When MD simulations are complete (this can take time), three remaining buttons should be run (create atom info files, create atom fluctuation files, parse/prepare files for DROIDS) should be run in that order. This runs a program for vector trajectory analysis called cpptraj and runs our custom perl script to prepare alignment files for later comparative dynamics analysis and visualization on the reference protein. When completed, a new GUI for statistical analysis will pop open (image below). We recommend using multiple test correction and strict significance cutoff (0.01) when running tests.





During statistical analysis, users are prompted to select chain and at the terminal to enter starting position of the chain. (default is zero) After statistics are run, 'goto DROIDS image and movie rendering' will open a new GUI that can control the resulting statistics as 'attribute files' that are readable by Chimera. Here, the statistics can be color-mapped to static or moving structures and saved as images in Chimera or viewed in our own movie viewer application. NOTE: movies are rendered on the first production MD run on the reference structure.





Images for ubiquitin cooled to 250K (top) showing dampened atom fluctuation and also run at same temperature (300K) (bottom) showing no significant change in fluctuation.

A simple DROIDS analysis would typically end here. Subroutine buttons for machine learning at the bottom of this GUI will convert the MD ensembles used for the simple DROIDS comparison into a training data set for machine learning that can be deployed on multiple new MD simulations for the purpose of analyzing where molecular dynamics are functionally conserved and the purpose of comparing genetic, epigenetic and drug class variant impacts on functionally conserved dynamics. This is explained fully in the subsequent maxDemon tutorials 5-8 below.

DROIDS TUTORIAL 2 – COMPARING ORTHOLOGS/PARALOGS AND CREATING MUTANTS

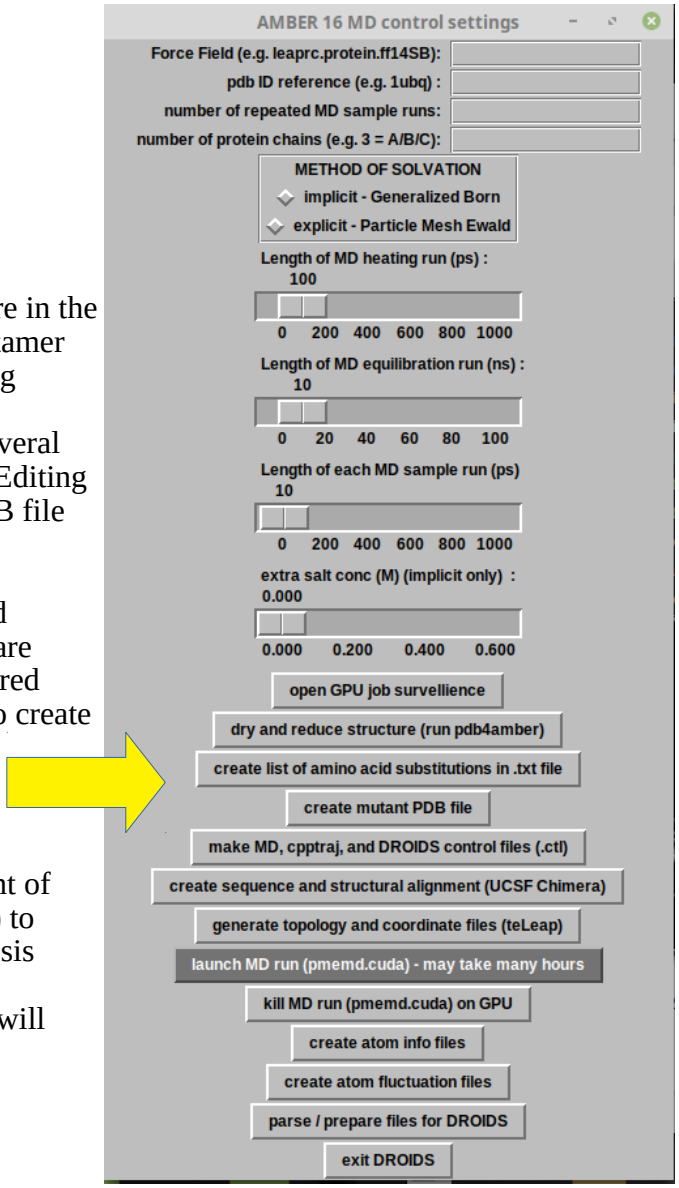
DROIDS offers pipelines for comparing pre-existing homologous PDB structures for simple proteins, and proteins bound to either DNA or small molecule ligand (main menu options # 3, 7 and 10) as well as pipelines that allow the user to create one or more mutations in protein or DNA (main menu options # 2, 6, and 9) from a list. The alignment step (see tutorial 1) is used to determine homology in the compared structures, and this is used to later resist DROIDS analysis and color mapping to occur on in the homologous regions. Non-homologous regions of the reference structure will be colored dark gray or tan in images and movies depending on the color schemes chosen. In menu option #3 (for simple protein homologs), the homology can be further defined as 'strict' or 'loose'. Strict homology requires that residues be identical at aligned sites, while loose homology allows then to be different.

When creating mutations, two additional GUI buttons are used, the first to write the list of mutations to a text file, the second to implement the list to alter the PDB file to create a mutant for comparison. Instruction for formatting the list is presented at the terminal and a text file with column headers is opened for this purpose. A list might look something like this.

substitution	position
ALA	23
TYR	31
PRO	35
LEU	47
ARG	52

This subroutine uses Chimera swapaa and swapna commands to create the structural variants, therefore in the case of amino acid modifications, the Dunbrack rotamer library is used to avoid steric clashes in the resulting mutant. If many amino acid replacements are implemented, we strongly recommend users run several thousand steps of Energy Minimization (Structure Editing tools in Chimera) on the mutant and resave the PDB file prior to loading into DROIDS.

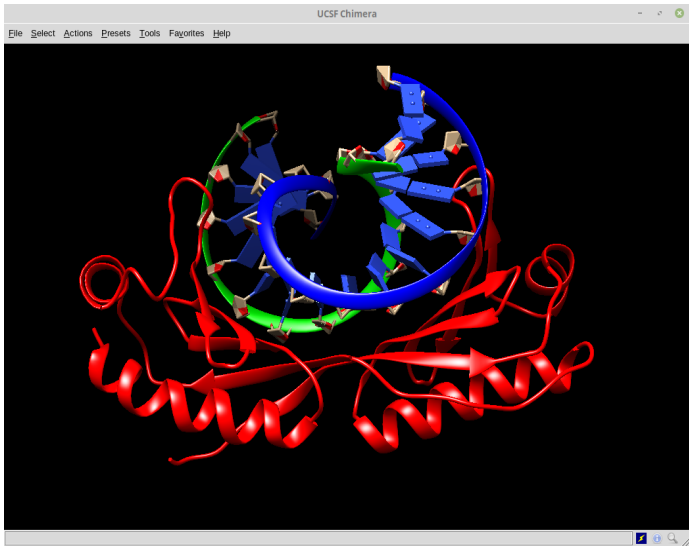
At this time, only base substitutions and amino acid replacements are supported. If insertion/deletions are required, the user can manually implement the desired changes using Structure Editing tools in Chimera to create homologs prior to running DROIDS.

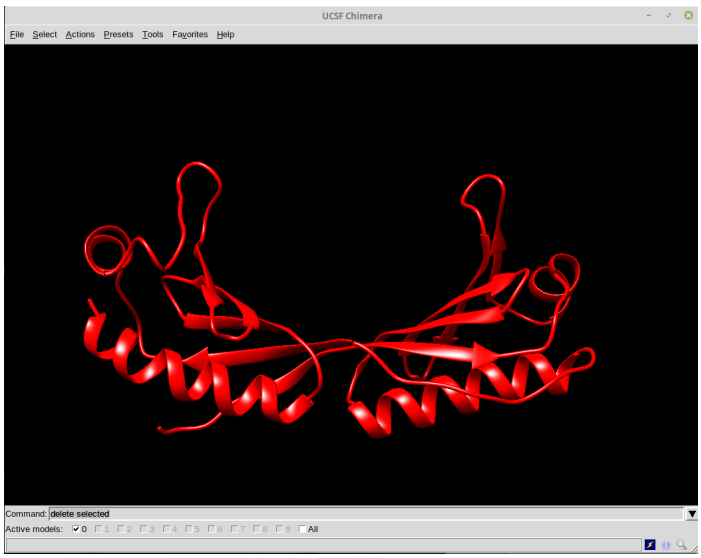


Mutant impacts can be analyzed from the standpoint of statistical significance (i.e. D values from KS tests) to predict potential results of a site-directed mutagenesis experiment, or in combination with drug binding interactions to to predict whether a genetic variant will cause changes in drug resistance or sensitivity.

DROIDS TUTORIAL 3 – ANALYZING MACROMOLECULAR BINDING INTERACTIONS

DROIDS offers pipelines for comparing proteins involved in binding interactions with DNA, other proteins, or small molecule ligands (i.e. drugs, toxins or signaling molecules like ATP). All of these are intended for comparing the protein dynamics in its bound state (query) to dynamics in its unbound form (reference). Visualizations of the statistics are presented on the bound structures. As these pipelines are very useful for defining the functional dynamic activity of proteins, they all have the maxDemon machine learning post-processing capacity for identifying functionally conserved dynamics as well as subsequent genetic or drug class variant impacts on these dynamics. To initiate these pipelines in DROIDS, users will need to create a bound and unbound starting structure, typically by deleting bound components and energy minimizing the structure. This is very easy to do in Chimera with editing commands and the resulting files can be saved with new PDB file names such as 1cdw_bound.pdb and 1cdw_unbound.pdb. The DROIDS GUI for DNA binding proteins will have an additional force field prompt for DNA as well as protein. A good modern force field for double-stranded DNA is leaprc.DNA.OL15. Single stranded nucleic acids are more problematic and can be loaded, however users should realize the the force fields are generally not intended to describe single strand dynamics, and MD can be very slow to sample all possible stable configurations in this state.





When working with DROIDS file preparation in the DNA bound structure, we recommend that DNA chains be labeled A and B and the subsequent protein chains start with C, D, E etc. Chimera allows users to relabel chains if needed. Another important step in this pipeline occurs during the creation of DROIDS control files. The .pdb files for structures are opened in the gedit text editor, and users are asked to delete the chain termination statements for DNA chains in the bound protein file. The line indicated by the blue arrow below is an example of what should be removed. There are typically two of these for a given DNA fragment. This is needed because later DROIDS scripts typically search for these termination statements when defining the lengths of chains of multichain proteins.

Open 1cdw.pdb ~/Desktop Save

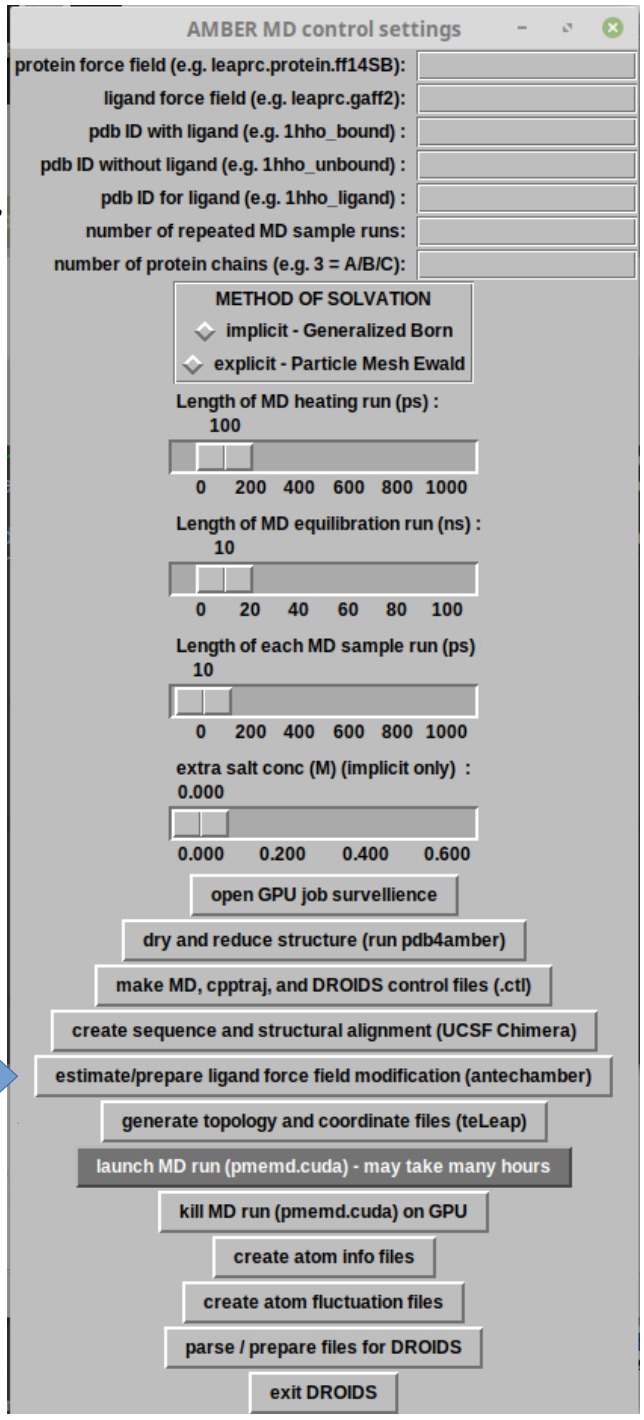
```
ATOM    311  C4'  DG B  16      68.603 119.672 -76.189  1.00 67.71           C
ATOM    312  O4'  DG B  16      69.779 120.371 -75.771  1.00 62.37           O
ATOM    313  C3'  DG B  16      67.627 120.741 -76.682  1.00 66.54           C
ATOM    314  O3'  DG B  16      66.803 121.164 -75.588  1.00 70.70           O
ATOM    315  C2'  DG B  16      68.589 121.853 -77.104  1.00 62.04           C
ATOM    316  C1'  DG B  16      69.600 121.768 -75.974  1.00 54.47           C
ATOM    317  N9   DG B  16      70.895 122.352 -76.315  1.00 45.32           N
ATOM    318  C8   DG B  16      71.610 122.276 -77.484  1.00 43.29           C
ATOM    319  N7   DG B  16      72.772 122.882 -77.432  1.00 42.67           N
ATOM    320  C5   DG B  16      72.821 123.393 -76.130  1.00 42.10           C
ATOM    321  C6   DG B  16      73.837 124.150 -75.476  1.00 37.83           C
ATOM    322  O6   DG B  16      74.923 124.500 -75.909  1.00 36.06           O
ATOM    323  N1   DG B  16      73.493 124.466 -74.179  1.00 35.35           N
ATOM    324  C2   DG B  16      72.311 124.098 -73.569  1.00 44.33           C
ATOM    325  N2   DG B  16      72.149 124.484 -72.304  1.00 48.89           N
ATOM    326  N3   DG B  16      71.348 123.384 -74.177  1.00 43.65           N
ATOM    327  C4   DG B  16      71.673 123.072 -75.450  1.00 42.07           C
TER     328       DG B  16
ATOM    329  O5'  DC C 101      78.809 128.266 -68.196  1.00 63.12           O
ATOM    330  C5'  DC C 101      77.873 128.811 -67.255  1.00 56.94           C
ATOM    331  C4'  DC C 101      76.417 128.530 -67.645  1.00 53.65           C
ATOM    332  O4'  DC C 101      76.223 128.626 -69.055  1.00 48.84           O
ATOM    333  C3'  DC C 101      76.019 127.103 -67.309  1.00 50.59           C
ATOM    334  O3'  DC C 101      75.582 127.016 -65.963  1.00 52.98           O
ATOM    335  C2'  DC C 101      74.828 126.877 -68.210  1.00 46.47           C
ATOM    336  C1'  DC C 101      75.164 127.733 -69.421  1.00 42.55           C
ATOM    337  N1   DC C 101      75.555 126.908 -70.563  1.00 36.97           N
ATOM    338  C2   DC C 101      74.557 126.180 -71.191  1.00 32.70           C
ATOM    339  O2   DC C 101      73.424 126.149 -70.722  1.00 36.33           O
ATOM    340  N3   DC C 101      74.861 125.492 -72.314  1.00 32.03           N
ATOM    341  C4   DC C 101      76.112 125.511 -72.796  1.00 34.64           C
ATOM    342  N4   DC C 101      76.374 124.855 -73.923  1.00 35.95           N
ATOM    343  C5   DC C 101      77.161 126.249 -72.147  1.00 34.39           C
ATOM    344  C6   DC C 101      76.836 126.926 -71.034  1.00 33.11           C
ATOM    345  P    DA C 102      75.833 125.645 -65.176  1.00 59.51           P
ATOM    346  OP1  DA C 102      75.570 125.896 -63.734  1.00 60.73           O
ATOM    347  OP2  DA C 102      77.144 125.100 -65.613  1.00 58.93           O
ATOM    348  O5'  DA C 102      74.674 124.692 -65.746  1.00 54.34           O
ATOM    349  C5'  DA C 102      73.323 124.995 -65.412  1.00 57.11           C
ATOM    350  C4'  DA C 102      72.354 124.016 -66.038  1.00 59.15           C
```

Plain Text Tab Width: 8 Ln 1, Col 1 INS

DROIDS TUTORIAL 4 – ANALYZING SMALL MOLECULE LIGAND BINDING

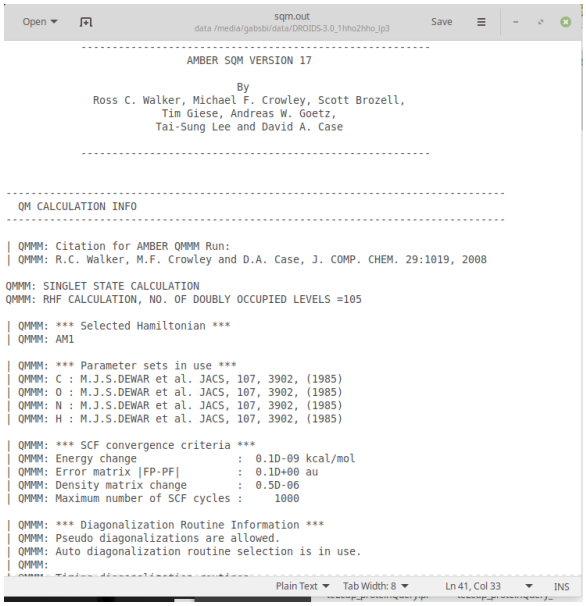
A major goal of DROIDS is to allow the biomedical research community to computationally investigate the functional dynamic impacts of protein interactions and small molecules, both natural and human designed. The force fields required for these sorts of simulations often require modification. Amber uses an automated command line pipeline for this process called Antechamber, which uses scaled quantum mechanical optimization routines to model the electron behavior on small molecule ligands and calculate modifications for the force fields that will allow the molecular dynamic simulations to proceed. Whenever a DROIDS pipeline for small molecules is requested on the main GUI, a button for running Antechamber will be present.

It is crucial that this process is run successfully prior to the setup of the rest of the system by teLeAP. In addition to the creation of a bound and unbound protein file, DROIDS will also require a 'ligand only' file, which can be easily created in Chimera in the same way the the unbound protein file...by deleting the protein chains and saving as a new pdb (e.g. 1yet_ligand.pdb). The MD GUI will also require a the loading of a general atom force field like gaff or gaff2, as well as this ligand file.



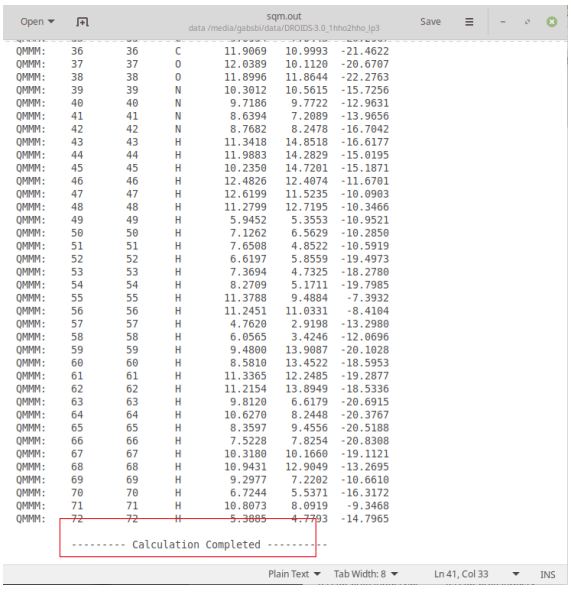
When antechamber runs, it may take some time. When the done, the DROIDS pipeline will attempt to open the sqm.out file, which allows the user to see results of the electron modeling. At the bottom of this text file will be an indication whether the modeling was successful or not. Directly after this is closed, DROIDS will attempt to open the .mol2 file in Chimera for the ligand that should have been created. NOTE; unlike .pdb format .mol2 format files will contain the newly estimated charge information produced by the sqm calculation. When the mol2 file in Chimera is closed, DROIDS will open the .frcmod file which should show all the force field modifications required by Amber. If these three files are not produced correctly, the user will not be able to proceed (and teLeAP will subsequently fail). Example files below are shown for the heme unit of hemoglobin after a successful pass through antechamber.

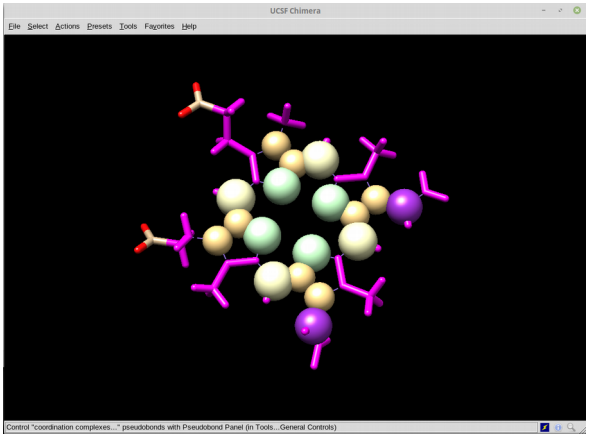
sqm.out (beginning of file)



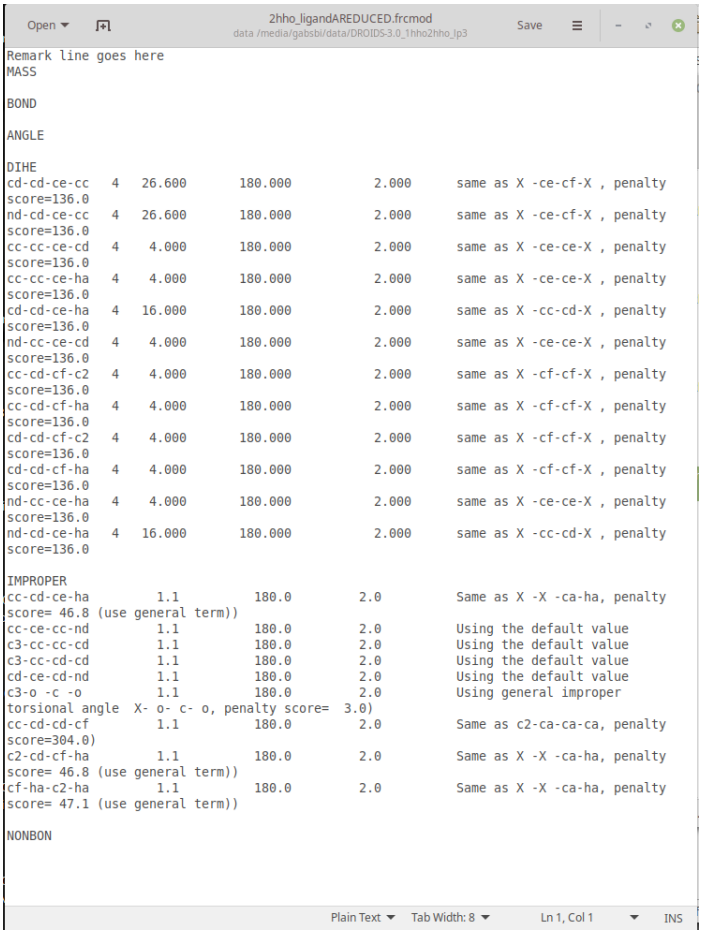
sqm.out (end of file)

make sure ‘calculation completed’ appears at bottom of this file



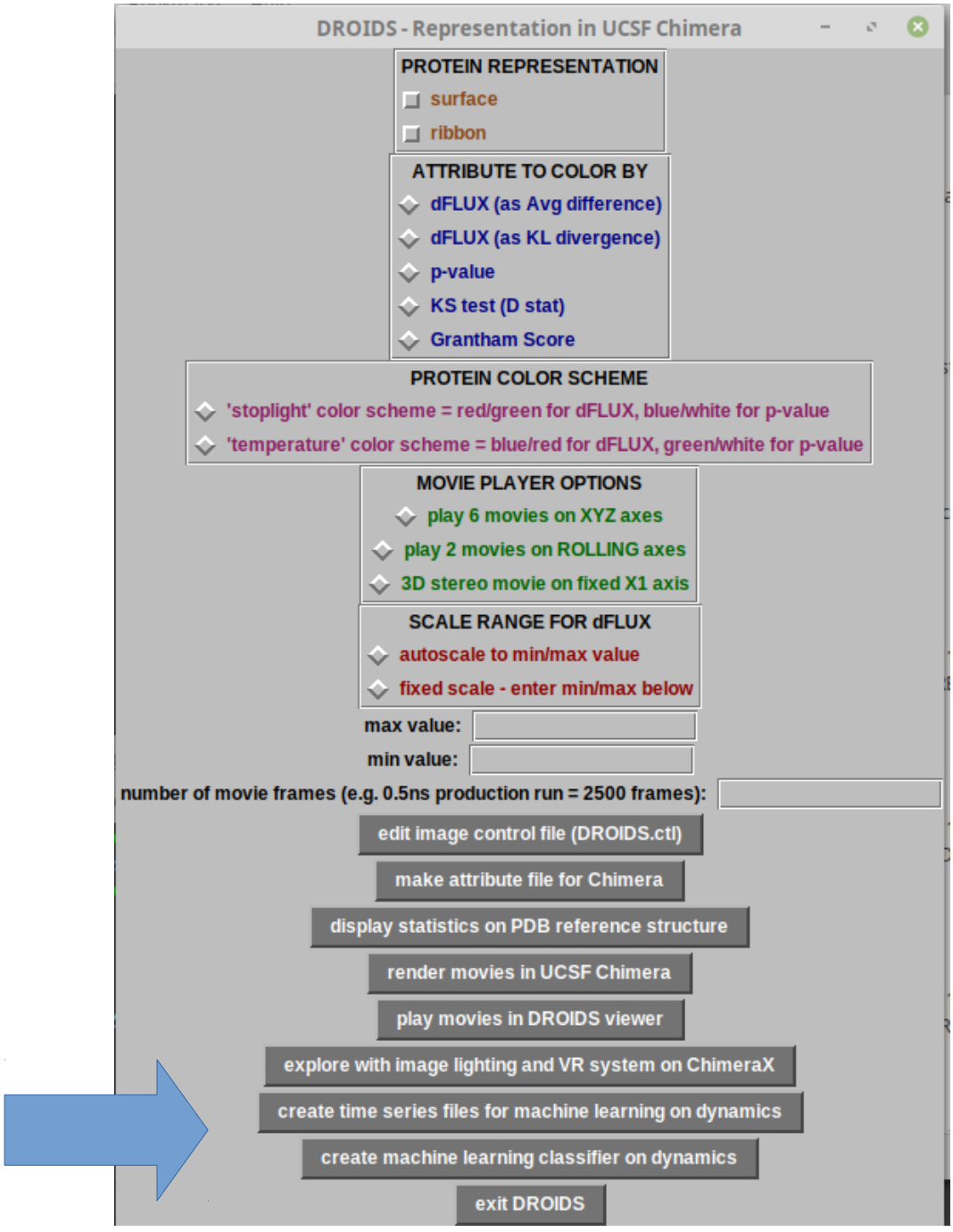


.mol2 structure + force field modification .frcmod file



MAXDEMON TUTORIAL 5 – DETECTING FUNCTIONALLY CONSERVED DYNAMICS

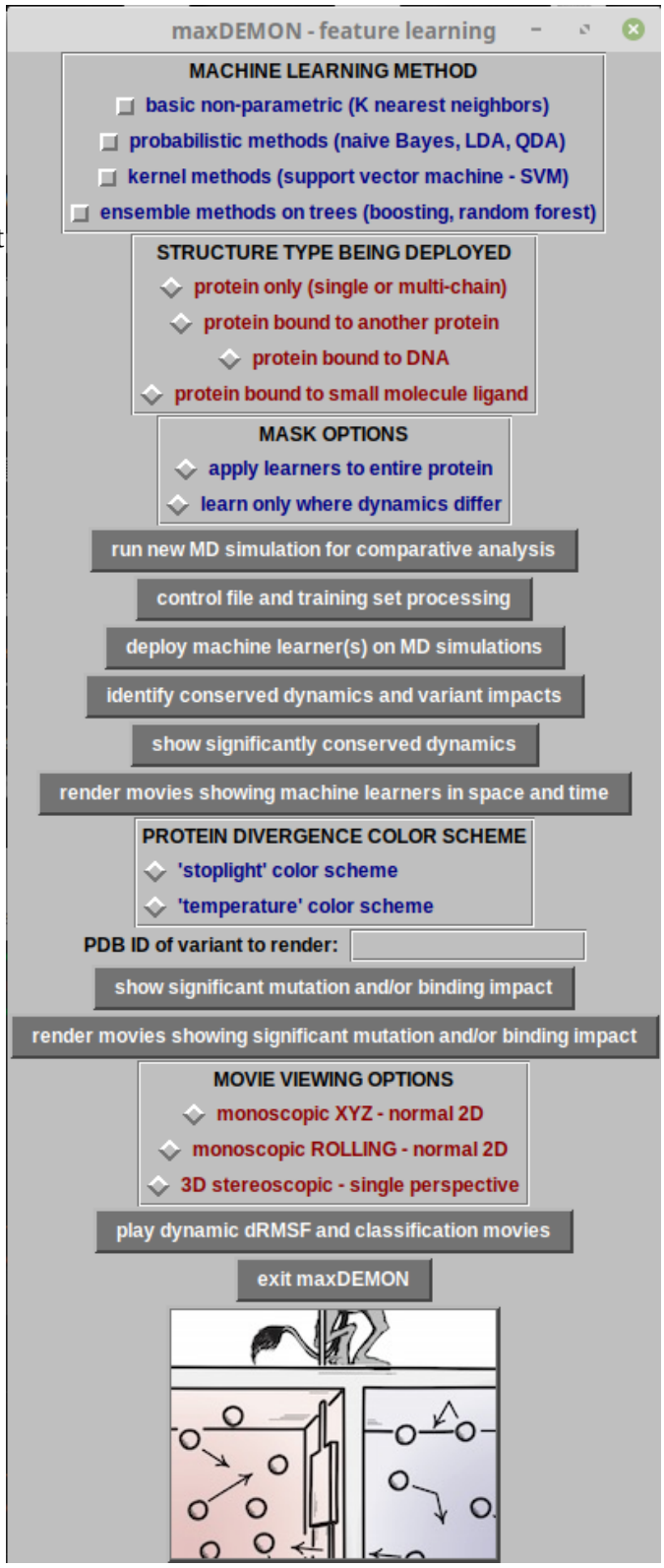
Whenever the user selects a DROIDS pipeline that represents a functional binding interaction (i.e. comparing protein in a bound vs. unbound dynamic state, a post-processing application maxDemon is available for launch at the bottom of the DROIDS image analysis GUI. As explained in our manuscript, maxDemon will use the functional comparison is DROIDS as a machine learning training set, allowing the resulting learning model to be applied to new molecular dynamic simulations. Two buttons here are available to first prepare time slices of the training sets, and second to launch maxDemon.



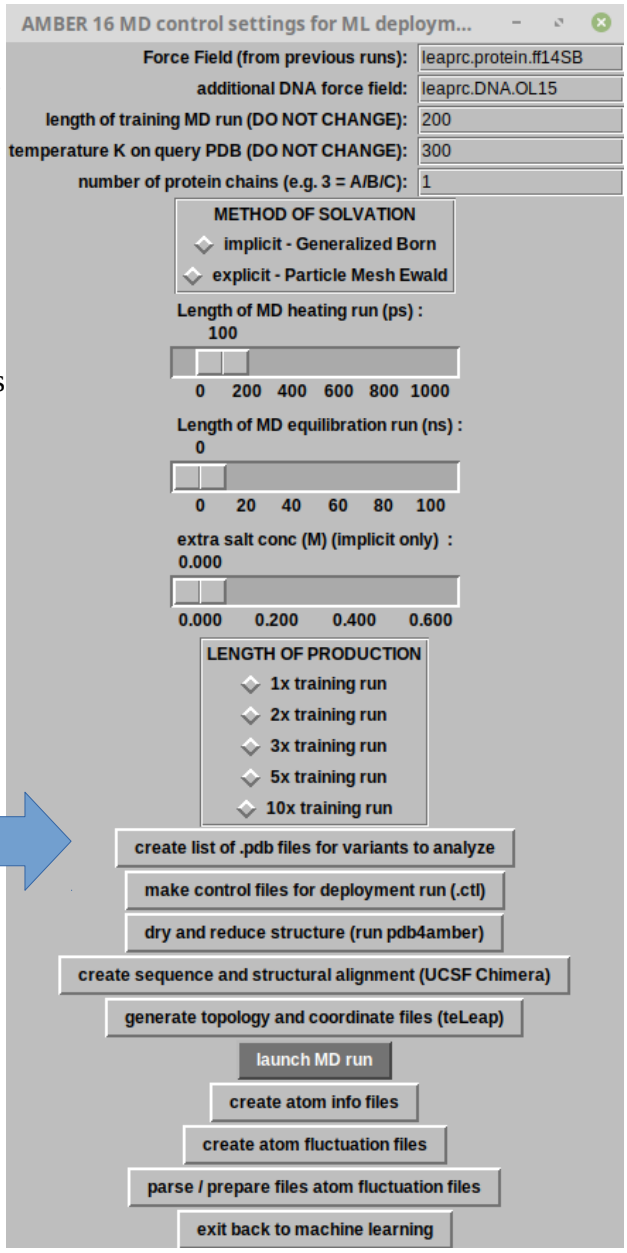
The first button/subroutine will run cpptraj commends to segment and recalculate the atom fluctuation values across time slices in the simulations that comprise the training set ensembles. The default time slice size is 50 frames(enetered at the terminal), however the user can increase the size of this time slice for longer runs. This step can take many minutes, depending upon the size of the system, size of ensemble, and length of the simulations.

The second button will open the main maxDemon software GUI shown here.

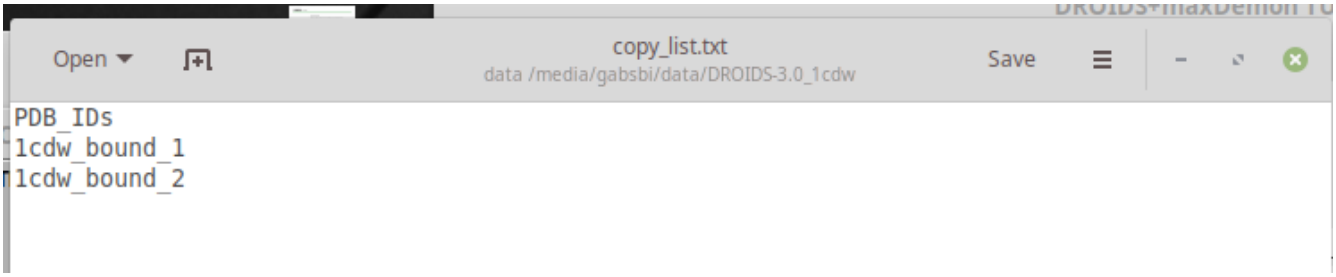
The steps involved at this stage are to define the 'stacked' learning model by selecting from the machine learning method menu at the top of the GUI, identify the structure type to be deployed (this should match the functional type of binding interaction upon which the training was done). Unless the system is very large, the leanring model should be applied to the entire protein and not only the regions where dynamics differ according to the multiple test corrected KS tests performed earlier in DROIDS. When the first button on this GUI is clicked, it will open another GUI designed for setting up and running additional molecular dynamics simulations upon which the learning model will be deployed. This will include at least two validation runs set up identically to the reference protein in the DROIDS training set. This list can be expanded to include drug class variants and genetic variants of interest.

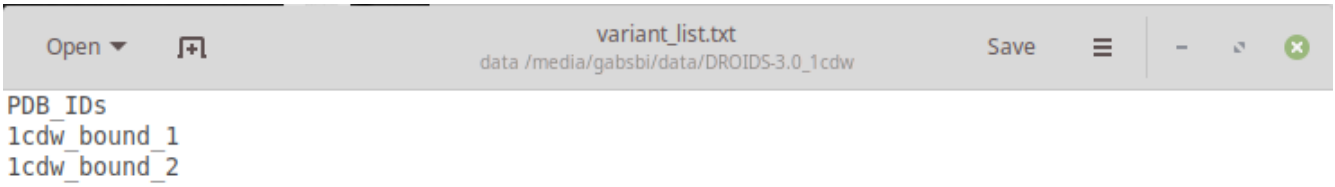


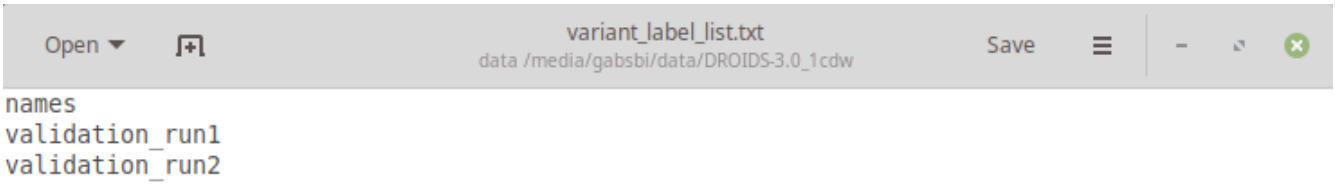
The Amber control GUI for setting up and launching these new simulations to which to apply machine learning is very similar to the control GUI used earlier in creating
the DROIDS training sets. The length of each of the production runs in the original molecular dynamic ensembles is typically very short, so there is an additional option to expand the length of the new simulations beyond the length of a single training run (e.g. 2x, 3x etc.). When the dynamic simulations are completed and post-processed to collect atom fluctuation values, the last button will take the user back to the main maxDemon GUI (on previous page) and the user can commence the deployment of the learning model on the new simulations.



NOTE: Prior to running the dynamic simulations from this GUI, the user will be prompted to create a list for all the .pdb files they want to analyze (blue arrow). The validation sets are created automatically and will appear at the top of each list. Three lists will pop open. These are shown on the next page below. The first (copy_list.txt) is for the validation set alone and does not usually need modified. The second file (variant_list.txt) will start with the validation sets and can be manually extended to include the names of any variant PDB structures set up by the user. The third list (variant_label_list.txt) allows the user to alter the naming conventions used in the R graphics produced at the end stage of the maxDemon analysis.

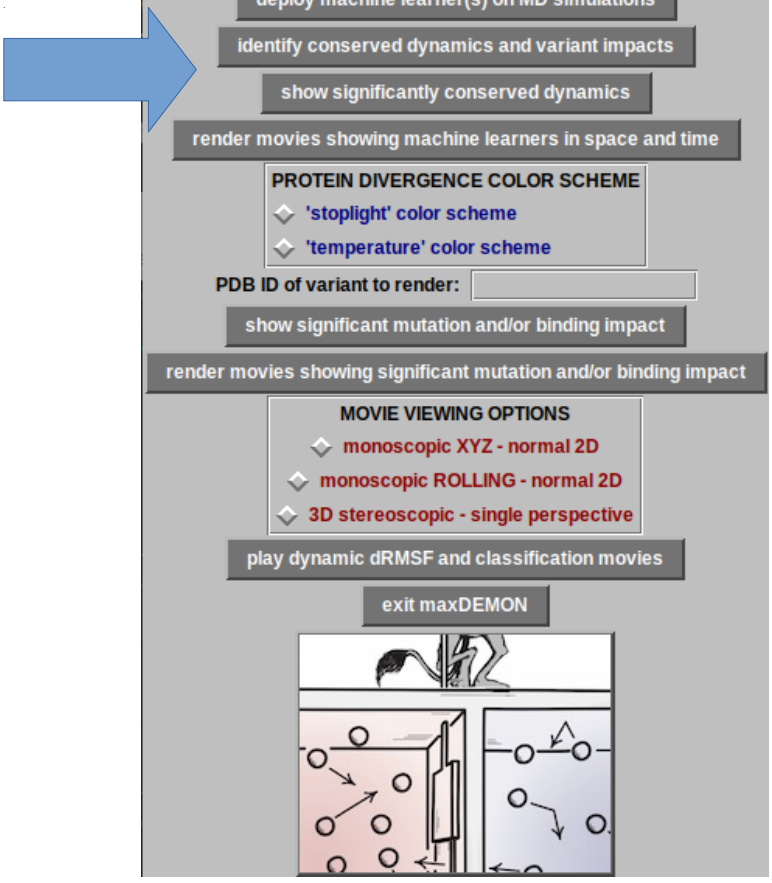




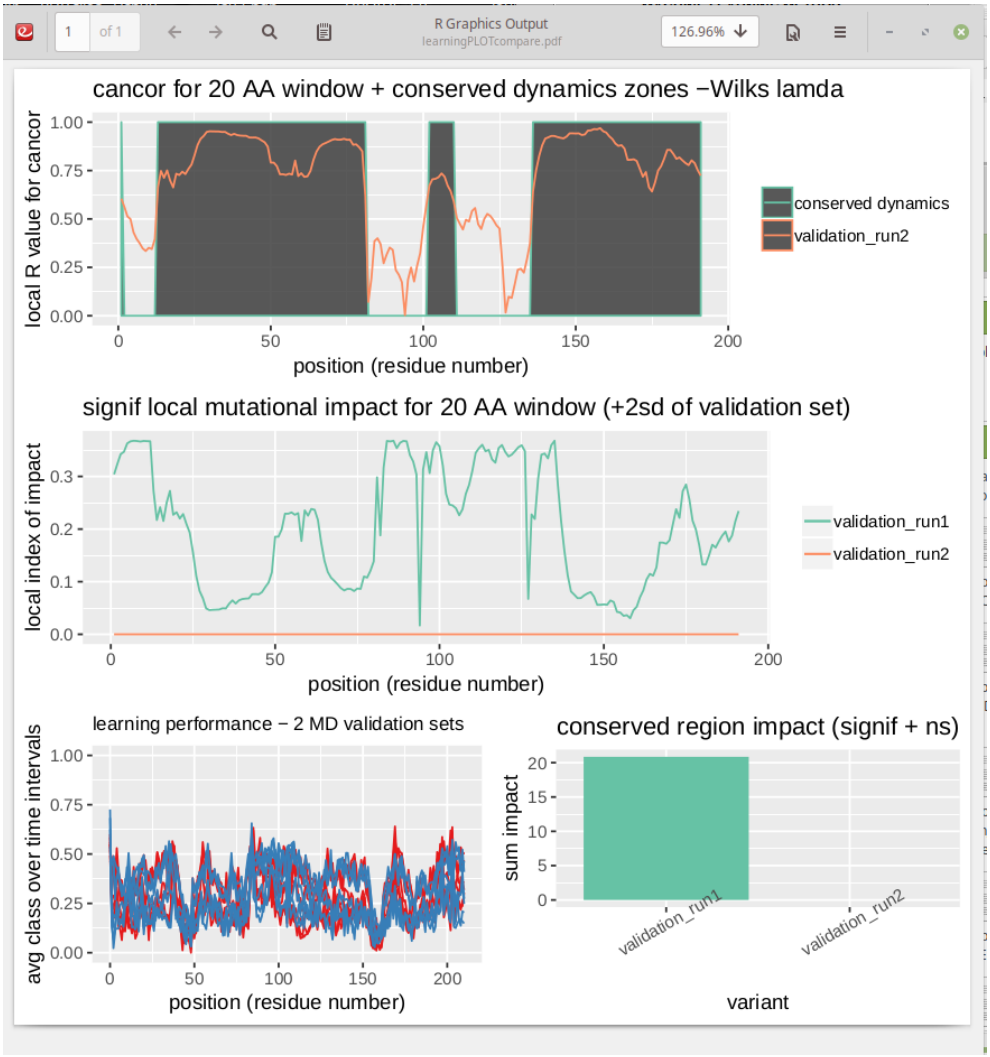


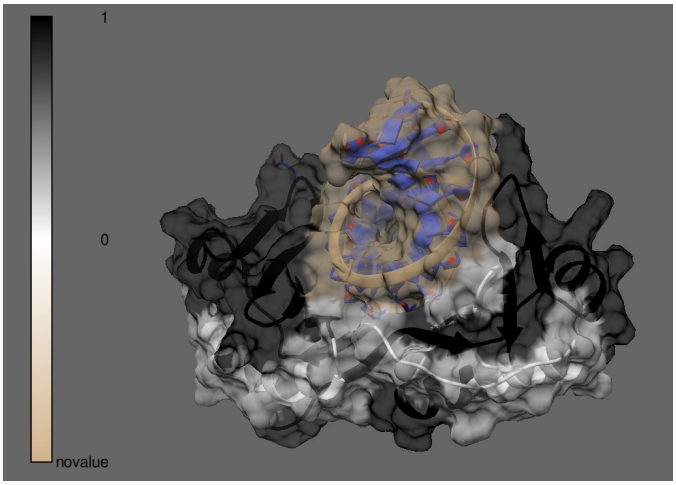
When the machine learning is deployed, the maxDemon R scripts will commence, and run the selected learning methods to every individual time slice for every amino acid residue in the structure. This can take some time to run. All the methods except KNN are programmed to use all available CPU cores on the machine. We generally recommend using all the learning methods when possible, but if run time is too long, the user can deselect the the SVM method or others to gain time. At least 3-4 learners should always be selected.

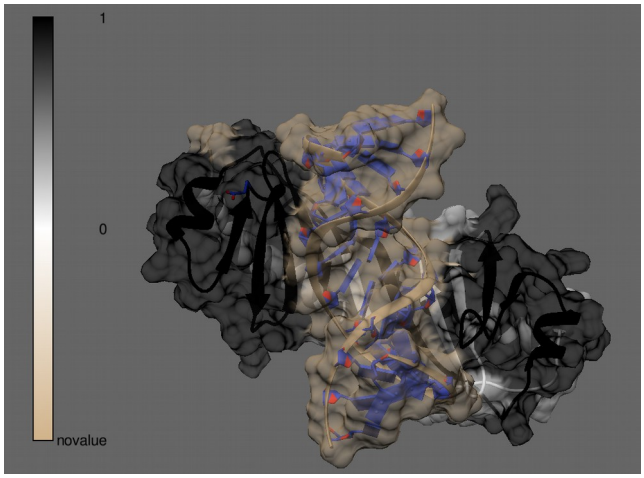
After the machine learning is completed, post-processing analysis of the results is controlled by the two buttons at the blue arrow below. The first will produce a composite R graphic analysis showing where the functionally conserved dynamics are on the protein (derived from validation runs) and the variant impacts on the protein dynamics (derived from the user supplied list). The user can chose to calculate impacts on dynamics over the total protein, or just the conserved regions. NOTE: conserved dynamic regions are shown in black on both the R plots and the PDB structure file.



Examples for an analysis of conserved dynamics in TATA binding protein (Tutorial #3) are shown below.



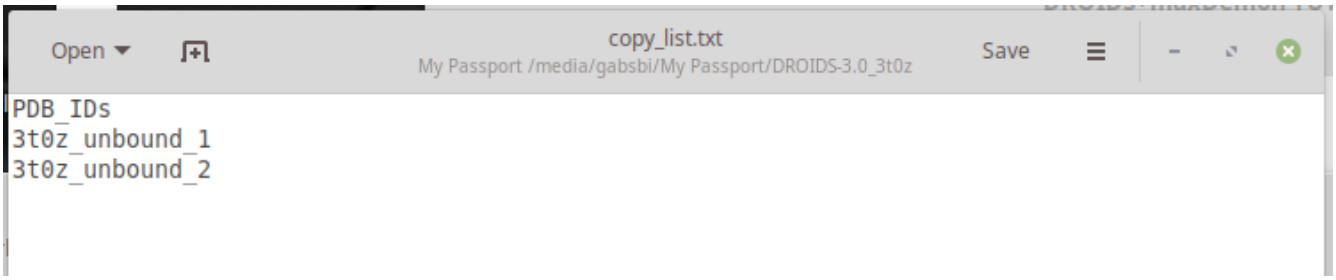




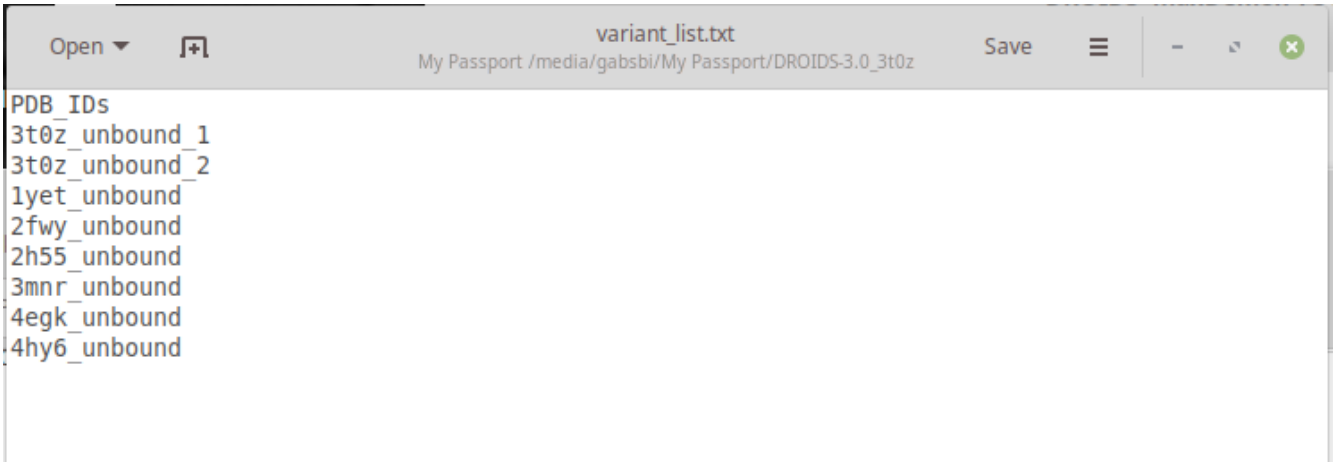
MAXDEMON TUTORIAL 6 – COMPARING DYNAMIC IMPACTS OF DRUG CLASS VARIANTS ON TARGET PROTEIN FUNCTION

Below are examples of a maxDemon analysis of drug class variants (PDB ID's 1yet, …) targeting the ATP binding pocket of Hsp90. The three lists used during setup are also shown. Aside from the addition of variant ID's and labels to the variant_list.txt and variant_label_list.txt, the analysis was conducted using steps outlined in Tutorial # 5. NOTE: as in previous DROIDS preps with ligands, the user must create a bound, unbound and ligand only .pdb file for each variant to be analyzed.

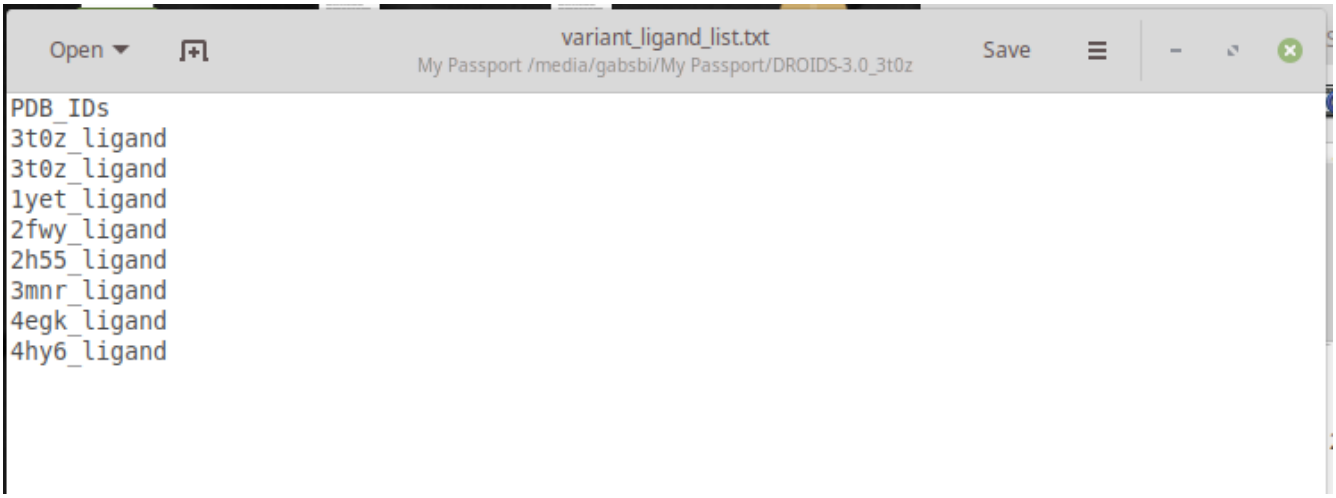
Two validation runs are requested at the terminal and this copy_list.txt is automatically generated and closed. NOTE: unbound structures are used because ligands will later be added from 'ligand only' after prep bt antechamber. Simply close it.



A new list will open (variant_list.txt) and the user will add the unbound variant file names as in example below. Then the file is saved and closed.

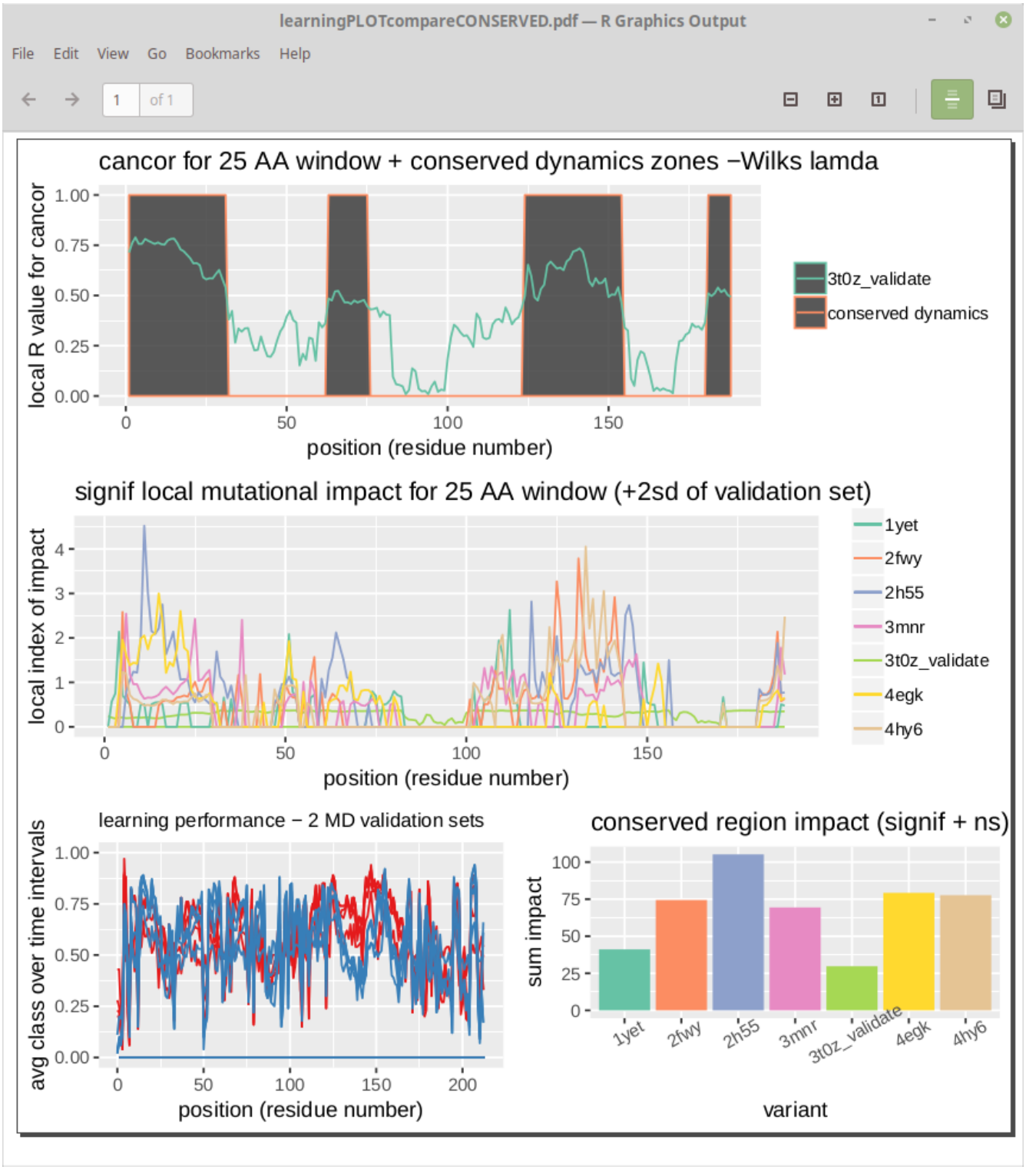


A third list is generated and saved for the names of all the ‘ligand only’ files



In a fourth list, the user can type labels for variants to be used in plots (use underscores, not blank whitespace)

The final analysis step will produce the multi-panel plot below. The regions of conserved dynamics are shown on the top plot, the variant impacts (calculated in a user defined sliding window) are shown in the middle plot, and the total impact (over whole protein OR conserved regions) is shown in the bar plot (bottom right). The stacked model machine learning performance during the two validation runs which is used to generate the canonical correlation from the top plot, are also shown (bottom left).



ADDITIONAL TIPS:

CREATING GENETIC MUTANTS IN UCSF CHIMERA FOR ANALYSIS IN MAXDEMON

Genetic variants can easily be created in Chimera prior to DROIDS+maxDemon analyses using the swapaa and swapna comands on the Chimera command line. Although swapaa uses the Dunbrack rotamer library to avoid steric clashes in mutated structures, the resulting structures can often benefit by energy minimization (using Structure Editing in Chimera), prior to running maxDemon.

LOADING STRUCTURES WITH EXISTING METAL IONS

If your structures contain functionally crucial metal ions, the .bat file that opens during teLeAP should have the following line added to load the library for ions. The charges should be double checked at the appropriate HETATOM lines in the .pdb file for each ion and manually corrected if necessary.

load atomic_ions.lib

You should keep in mind that MD force fields may not correctly capture the behavior of bonded interactions with metals.

LOADING MULTIPLE LIGANDS OR LIGANDS WITH MULTIPLE PARTS

Antechamber/sqm software in Amber requires that small molecule ligand files consist of a single unbroken structure with at least 1 chemical bond. Our GUIs for protein ligand interactions only contain input for a single ligand file. If multiple ligands or multiple parts of a single ligand are needed to be setup, the user can make multiple ligand files, each with a different name (ligandA, ligandB, ligandC or partA, partB and partC) and enter them each one at a time, subsequently running the antchamber subroutine. When creating the solvated system of the bound structure in teLeAP, additional lines should be manually added to the .bat file to load the additional ligands or parts of ligands.

LOADING SYSTEMS WITHIN LIPID BOUNDARIES

Force fields for lipids do exist in the Amber library. Systems with membrane bound components could be analyzed with proper modification of the .bat files popped open during teLeAP. Additional lines to load the lipid force field and edits to properly size the water box would be needed at minimum.

LOADING SYSTEMS WITH PRE-PREPARED PROTONATION STATES

When running the ‘dry and reduce’ button on the Amber MD GUI, there is a terminal prompt that asks ‘y’ to reduce the whole structure and ‘n’ to skip this step. If skipped the program will run pdb4amber without the –dry option. This will rewrite the pdb file, removing crystallographic waters, without altering hydrogens. If a prefered protonation state is set up manually in the pdb file prior to running DROIDS it can be maintained with this option. NOTE: teleap will expect residues like histidine with specific protonation states to be properly specified in the pdb file (e.g. HIE or HID) or else unspecified (e.g. HIS) so some manual curation of residue labels in the pdb file can prevent later problems in teLeAP setup for MD.

LOADING SYSTEMS WITH ELEVATED SALT CONCENTRATIONS

DROIDS is programmed to set up simple charge neutralized systems using Na+ and Cl- ions added automatically during teLeAP setup. We recommend the following method/software for

estimating number of ions to add when constructing systems with elevated salt conditions. The .bat file lines to add Na+ and Cl- ions end in zero (to specify automatic charge neutralization) and should be rewritten to specify the exact number the user wants to add.

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.

SHOULD I USE IMPLICIT OR EXPLICIT SOLVATION?

While DROIDS and Amber pmemd.cuda can employ implicit solvation under Generalized Born method, we recommend against it. It generally will not replicate statistically insignificant differences in dynamics on identical structures. Most experienced MD users employ explicit solvation with the Particle Mesh Ewald method, so we recommend most users aiming for publishable results do the same. We suspect that our implicit solvent setup in DROIDS is simulating protein behavior in more-or-less conditions of a vacuum. The proper parameter settings should be added in .bat file to simulate continuum applied by a specific solvent condition. Experienced Amber users can play around with this, but we recommend most biological applications use explicit solvation with the most efficient water model Tip3p, or more realistic water models (Tip4p, Tip5p and Tip6p).

See our USER MANUAL 3.0 for additional tips in using DROIDS