MOSAICA

(MOSAICS from Mathematica)

Alberto Hernández-Espinosa
Hector Zenil
July, 2015
University of Oxford.

MOSAICA is a Mathematica package for running into an integration environment a set of biology computational analysis tools.

Description of tools:

MOSIACS: is a software package designed to refine molecular conformations directly against two-dimensional (2D) electron-microscopy images. By optimizing the orientation of the projection at the same time as the conformation, it is particularly well- suited to the 2D class-averages from cryo-electron microscopy. By directly using projection images, we relieve the urgent need for a density map that is not always available due to the structural heterogeneity or preferred orientations of the sample molecules on the grid.
Source: http://www.cs.ox.ac.uk/mosaics/documentation/mosaics_em_guide.pdf

pymol: user-sponsored molecular visualization system on an open-source foundation.
Source: https://www.pymol.org/

x3dna: Suite of programs to analyze, rebuild, and visualize three-dimensional nucleic-acid structures. The software determines a wide range of conformational parameters, including the identities and rigid-body parameters of interacting bases and base-pair steps, the nucleotides comprising helical fragments, the area of overlap of stacked bases, *etc*.
Source: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3065354/

Curves+: A revised version of the Curves approach for analysing the structure of nucleic acids. It respects the international conventions for nucleic acid analysis, runs much faster and provides new data.
Source: https://bisi.ibcp.fr/tools/curves_plus/

MOSAICA: Due above software is implemented in differente technologies, even they seems to work in a common line of investigation, we think it would be helpful to have an integral invironment to run all these tools in order to agilize sharing information between this set of biological anaylisis tools.
Mathematica as a language let us to do this.
To make MOSAICA works it is necesary to set a complete development environmet. In the next pages is show how to do this.
User has to be advised that detailed instalation instructions are supplied in software packages sites.

SECTION 1
Downloading and installing software

1.1 Downloading and installing MOSAICS

a) Download software from site: http://www.cs.ox.ac.uk/mosaics/
Site will ask for simple registration. After submit package is dowloaded.
In this version of MOSAICA, it is used MOSAICS.3.8-EM.

b) installing MOSAICS
Details of installation can be found in
http://www.cs.ox.ac.uk/mosaics/install/installation_mosaics_em.html

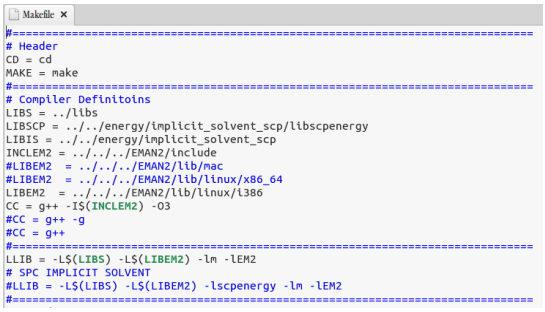
In the next images is shown how MOSAIC.3.8-EM was installed in a core i7 running under UBUNTU 14.04, 64 bits.

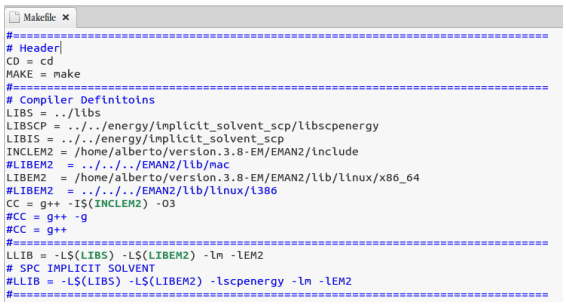
b.1. Unpacking

$ tar -xvzf version.3.8-EM.tar.gzD

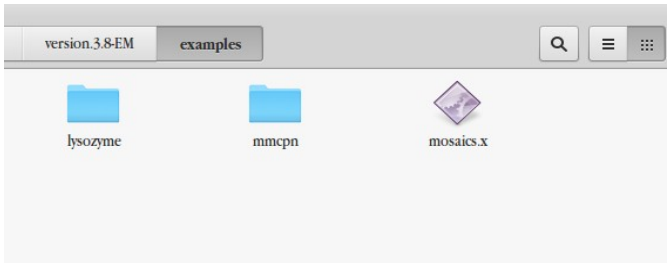
b.2 Compilling

b.2.1. cd source/compile/serial
b.2.2. Edit Makefile. First, orginal Makefile. Second, Makefile after edition





b.2.3. $ make clean
b.2.4. $ make



b.3. Setting environment variables: the LD_LIBRARY_PATH environment variable must point to libME2.so path. For example:

$ export LD_LIBRARY_PATH=/home/alberto/version.3.8-M/EMAN2/lib/linux/x86_64:$LD_LIBRARY_PATH

b.4 Running test:

cd version.3.8-EM/examples/lysozyme

Creation links to executables and libraries:

:~/version.3.8-EM/examples/lysozyme$ ln -s /home/alberto/version.3.8-EM/EMAN2/lib/linux/x86_64/libEM2.so libEM2.so

:~/version.3.8-EM/examples/lysozyme$ ln -s /home/alberto/version.3.8-EM/examples/mosaics.x mosaics.x

Runing test:
:~/version.3.8-EM/examples/lysozyme$ ./mosaics.x refine.input>out

1.2. Installing x3DNA (from: http://forum.x3dna.org/howtos/how-to-install-3dna-on-linux-and-windows/)

Assuming the downloaded tarball file is named x3dna-v2.1-linux-64bit-yyyymondd.tar.gz (where yyyymondd is the date of the release, e.g., 2012jun26). At your current working directory (presumably your home directory), do the following:

- tar pzxvf x3dna-v2.1-linux-64bit-yyyymondd.tar.gz
  Here the options pzxvf require some explanation:
  - **p** to preserve permissions of the various directories and files. This option may not be required for a personal setting.
  - **z** filter the archive through gzip. With this option, we can work directly from .tar.gz file without first using gunzip.
  - **x** to extract files from an archive.
  - **v** to verbosely list files processed.
  - **f** to use archive file, i.e., the 3DNA tarball file x3dna-v2.1-linux-64bit-yyyymondd.tar.gz.

  After running the above command, you will get a directory named x3dna_v2.1 which contains the 3DNA v2.1 distribution.
- cd x3dna_v2.1/bin
  enter into the bin directory of the 3DNA v2.1 distribution.
- ./x3dna_setup
  run the 3DNA setup Ruby script, and you will see an output similar to the following:
  To install X3DNA, do as follows:
  (0) download 3DNA binary distribution for your system from URL
  http://x3dna.org
  (1) tar pzxvf x3dna-v2.1-linux-64bit-yyyymondd.tar.gz
  (2) cd x3dna-v2.1/bin
  (3) ./x3dna_setup
  To run X3DNA, you need to set up the followings:
  o the environment variable X3DNA
  o add $X3DNA/bin to your command line search path
  for your 'bash' shell, please add the following into ~/.bashrc:
  ---
  export X3DNA='/home/xiangjun/x3dna-v2.1'
  export PATH='/home/xiangjun/x3dna-v2.1/bin':$PATH
  ---
  and then run: source ~/.bashrc
  (4) type find_pair -h
  for command line help -- this applies to all 3DNA binaries
  Visit 3DNA homepage at URL http://x3dna.org/ for more info.
  Here the key parts are colored red, i.e., set the environment variable X3DNA and add $X3DNA/bin to your command search path.

The above example refers to the most commonly used 'bash' shell. For the 'sh' shell, the 'x3dna_setup' output would be:

for your 'sh' shell, add the following into ~/.profile:

---

export X3DNA='/home/xiangjun/x3dna-v2.1'
export PATH='/home/xiangjun/x3dna-v2.1/bin':$PATH

---

and then run: source ~/.profile

You *may* need to use the **dot** command: . ~/.profile. For the 'tcsh' (or 'csh') shell, the 'x3dna_setup' output would be:

for your 'tcsh' shell, add the following into ~/.cshrc:

---

setenv X3DNA '/home/xiangjun/x3dna-v2.1'
setenv PATH '/home/xiangjun/x3dna-v2.1/bin':$PATH

---

and then run: source ~/.cshrc

If your shell is none of the four recognized ones -- bash, sh, tcsh, or csh -- or cannot be detected, the script still outputs sensible settings about the installation.

- find_pair -h

  or ./find_pair -h if you've strictly followed the steps above, and your current working directly is not in the command search path. Now you should see a screenful of command-line help. Congratulations -- You now have 3DNA properly installed! If otherwise, please post here with details, and we will try to help you out.

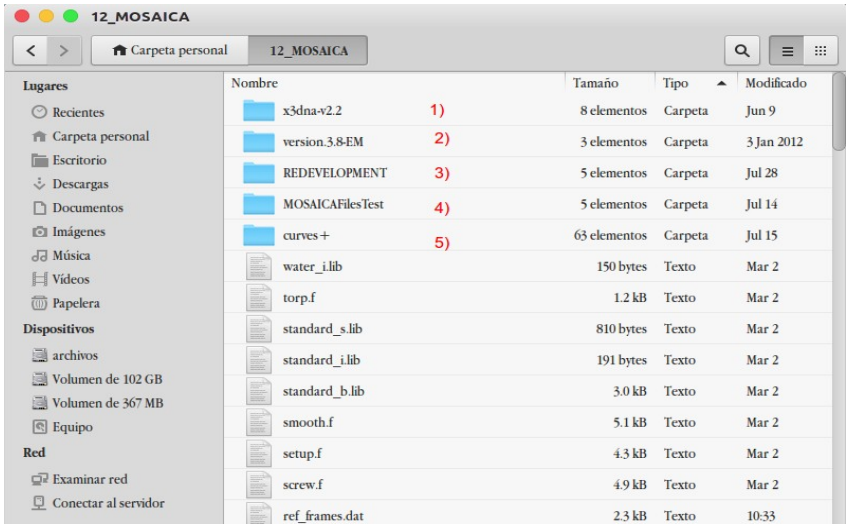
1.3. Installing Curves+

After download it, apply “make” command.
(source: https://bisi.ibcp.fr/tools/curves_plus/ewExternalFiles/Curves+new.pdf)

1.4. Installing pymol

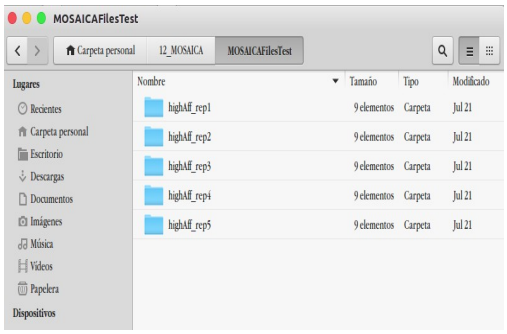
# Debian/Ubuntu/Mint
sudo apt-get install pymol

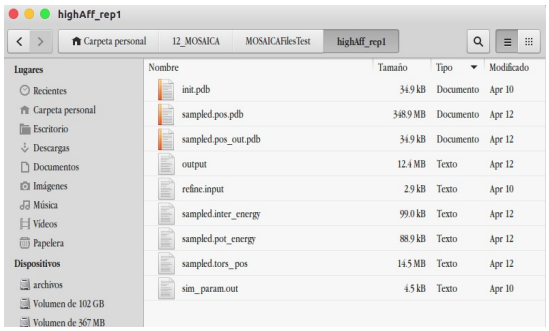
2. Setting MOSAICA environment.



SECTION 2: MOSAICA ENVIRONMENT

1) x3dna installation folder
2) MOSAICS installation folder
3) REDEVELOPMENT folder: where all results will be hold. At the beggining, this folder must be empty. After running MOSAICA, this will contain all results.
4) MOSAICAFilesTest Folder: Contains results of MOSAICS simulations. One simulation per subfolder. In this way:

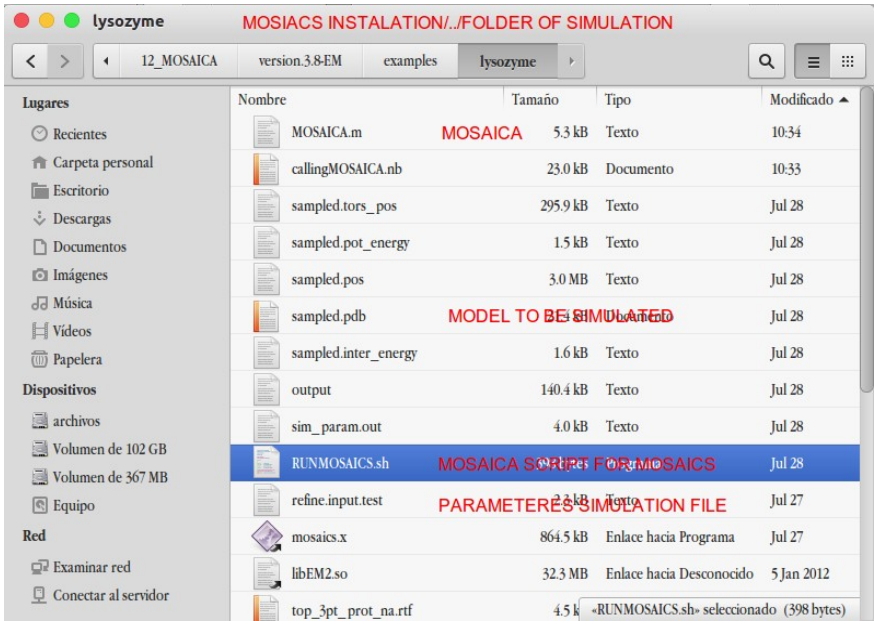




5) Curves+ installation folder: Content of this folder must be pasted in the MOSAICA folder.

6) Others: Since the way MOSAICA "control" the analisys environment is by means scripts od different nature, these scripts must be contained in MOSAICA folder:

Note: When a MOSAICS simulation runs, MOSAICA.m and RUNMOSIACS.sh, simluation parameters file and model (.pdb) files must exists in the simulation folder:



SECTION 3: HOT TO USE MOSAICA

0) Specification of simulation pararameters
1) Run MOSAICS simulation
2) Save any simulation in a exclusive folder
3) Run x3dna analisys over all simulations
4) Run Curves+ analisys over all simulations

How to see and specify simulation parameters

ShowSimulationParamValues function show current parameters values

```
{$Context, $ContextPath};
Names["MOSAICA`"];
Names["Global`"];
ClearAll["MOSAICA`"];
ClearAll["Global`"];
SetDirectory["/home/alberto/12_MOSAICA"];
Get["MOSAICA`"];
SetMOSAICADirectory["/home/alberto/12_MOSAICA"];
SetSimulationPath["/home/alberto/12_MOSAICA/version.3.8-EM/examples/lysozyme"];
SetSimulationParametersFile["refine.input.test"];
ShowSimulationParamValues
```

```
simulation_typ{MIN}  PT EEMC SEQ_PT SEQ_EEMC NM DBFR
minimize_type{stsamc}
prop_tors_sig{1.e-2} 1.e-5 proposal sig 0 < number < 2 Pi, usually 1.e-5
prop_rot_sig{1.e-4} 1.e-5 {0 <= radian < 2 Pi}
prop_trans_sig{1.e-3} 1.e-4 {Angstrom >= 0.0}
prop_clos_sig{1.e-3} {Angstrom >= 0.0}
total_step_mc{7000} 2000000 :10   number > 1
statistics_freq{100} 200
write_energy_unit{Ha} kcal  Ha: atomic unit, kcal: kcal/mol
stsamc_type{trigonom}
stsamc_period{4000}
stsamc_ampl{2500}
stsamc_shift{0}
random_seed{-9378000501} {large integer}
cgres_model{KB_3pt}  KB_3pt, off
mol_parm_file{top_3pt_prot_na.rtf}
inter_database_file{par_3pt_prot_na.prm}
cryo_em_database_file{orientation.data}
pos_init_file{init.pdb}
pos_out_file{sampled.pdb}
atom_pos_file{sampled.pos}
epot_file{sampled.pot_energy}
einter_file{sampled.inter_energy}
energy_term{bond_off}
energy_term{bond_off}
```

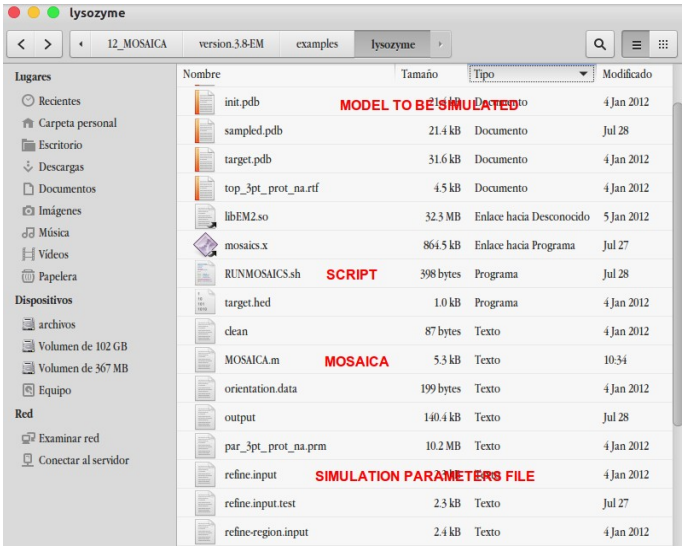
Changing a simulation parameter value

ReplaceSimulationParameter[“name_of_parameter”, “new_value”]

```
(*Remove[SetMOSAICADirectory, SetSimulationParametersFile, ReplaceSimulationParameter, myFile]*)
{$Context, $ContextPath}
Names["MOSAICA`"]
Names["Global`"]
ClearAll["MOSAICA`"]
ClearAll["Global`"]
SetDirectory["/home/alberto/12_MOSAICA"];
Get["MOSAICA`"]
SetMOSAICADirectory["/home/alberto/12_MOSAICA"];
SetSimulationPath["/home/alberto/12_MOSAICA/version.3.8-EM/examples/lysozyme"];
SetSimulationParametersFile["refine.input.test"];
#ReplaceSimulationParameter["num_procs", "ghyt"]
```

Runing MOSAICS simulation

In simulation folder must contain model file (.pdb), MOSAICA scritp, MOSAICA.m and simulation parameters file.



RunMOSAICSSimulation function execute MOSAICS simulation

```
{$Context, $ContextPath};
Names["MOSAICA`"];
Names["Global`"];
ClearAll["MOSAICA`"];
ClearAll["Global`"];
SetDirectory["/home/alberto/12_MOSAICA/version.3.8-EM/examples/lysozyme"];
Get["MOSAICA`"];
SetMOSAICADirectory["/home/alberto/12_MOSAICA/version.3.8-EM/examples/lysozyme"];
SetSimulationPath["/home/alberto/12_MOSAICA/version.3.8-EM/examples/lysozyme"];
SetSimulationParametersFile["refine.input.test"];
RunMOSAICSSimulation
```

After every simulation, and before running new simulations, results must be saved in a folder. Every serie of simulations must be saved in the same folder in the master MOSAICA directory.

/home/alberto/12_MOSAICA //Master folder, MOSAICA
/home/alberto/12_MOSAICA/MOSAICAFilesTest //container of serie simulations
/home/alberto/12_MOSAICA/MOSAICAFilesTest/highAff_rep1 //--> simulation 1
/home/alberto/12_MOSAICA/MOSAICAFilesTest/highAff_rep1/init.pdb // result of simulation 1
/home/alberto/12_MOSAICA/MOSAICAFilesTest/highAff_rep1/sampled.pos.pdb // result of simulation 1
/home/alberto/12_MOSAICA/MOSAICAFilesTest/highAff_rep1/sampled.pos_out.pdb // result of simulation 1
/home/alberto/12_MOSAICA/MOSAICAFilesTest/highAff_rep2 //--> simulation 2
… //--> serie of simulations
/home/alberto/12_MOSAICA/MOSAICAFilesTest/highAff_repn //--> simulationn

Reordering using pymol

An empty folder for analysis results is needed. In the next example MOSAICAFilesTest contains MOSAICS simulations, and REDEVELOPMENT is the empty folder.

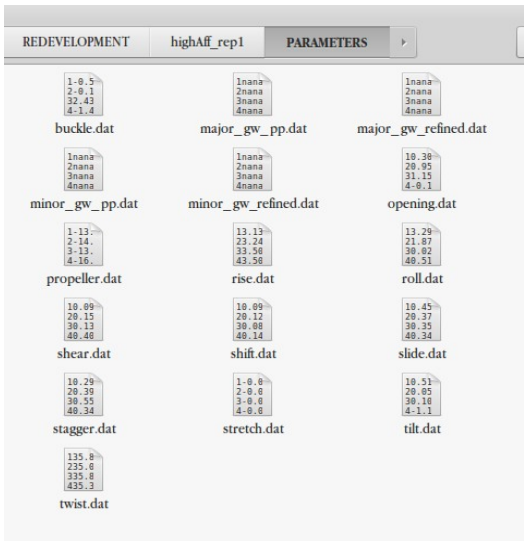
RunPymolReordering functions format MOSAICS results to be handled for further analisys.

```
{$Context, $ContextPath};
Names["MOSAICA`"];
Names["Global`"];
ClearAll["MOSAICA`"];
ClearAll["Global`"];
SetDirectory["/home/alberto/12_MOSAICA"]
Get["MOSAICA`"]
SetMOSAICAMasterPath["/home/alberto/12_MOSAICA"];
SetMOSAICSTestsDirectory["/home/alberto/12_MOSAICA/MOSAICAFilesTest"];
SetReorderingTestsDirectory["/home/alberto/12_MOSAICA/REDEVELOPMENT"];
RunPymolReordering
/home/alberto/12_MOSAICA
```

X3dna analisys.

RunX3DNAAnalisys function run x3dna on reodered files by pymol. This fucntion generates a new folder with all parameteres to be analyzed.

```
{$Context, $ContextPath};
Names["MOSAICA`"];
Names["Global`"];
ClearAll["MOSAICA`"];
ClearAll["Global`"];
SetDirectory["/home/alberto/12_MOSAICA"]
Get["MOSAICA`"];
SetMOSAICAMasterPath["/home/alberto/12_MOSAICA"];
SetReorderingTestsDirectory["/home/alberto/12_MOSAICA/REDEVELOPMENT"];
SetX3DNADirectory["/home/alberto/12_MOSAICA/x3dna-v2.2"];
RunX3DNAAnalisys
Out[6]= /home/alberto/12_MOSAICA
```



Curves Analisys.

RunCurvesAnalysys function create a bridge between x3dna and Curves+ software. This is to say. By mens x3dna functions, a bridge file is built on order to link pdb files and results of x3dna analysis in order to execute curves+ analysis.

```
{$Context, $ContextPath};
Names["MOSAICA`*"];
Names["Global`*"];
ClearAll["MOSAICA`*"];
ClearAll["Global`*"];
SetDirectory["/home/alberto/12_MOSAICA"]
Get["MOSAICA`"];
SetMOSAICAMasterPath["/home/alberto/12_MOSAICA"];
SetCurvesPath["/home/alberto/12_MOSAICA"];
SetReorderingTestsDirectory["/home/alberto/12_MOSAICA/REDEVELOPMENT"]
SetX3DNADirectory["/home/alberto/12_MOSAICA/x3dna-v2.2"];
RunCurvesAnalysys
```

Out[176]= /home/alberto/12_MOSAICA

Out[180]= /home/alberto/12_MOSAICA/REDEVELOPMENT