IGEM IONIS

RNA Modeling

Simulation and Analysis

with

SimRNA, GROMACS, and VMD

Gaetan Christien, Nicolas Delettre, Thomas Lhernould
Advisor: Alexandre Ismail, Ph.D
IONIS-PARIS 2017 | IONIS.IGEM@gmail.com

Table of contents

Introduction

This document has been written to facilitate the use of modelling software by students and researchers in biology. The writing of this document has been motivated by the will to gather practical information in one document. Our handbook describes how to install and run software to simulate and visualize RNA folding. We will not go deep into the details of how the software works. There are alternatives to our software choices.

We are three students from Sup'Biotech, a biotechnology engineering school just outside of Paris, France, in Villejuif. We are part of the iGEM IONIS-PARIS 2017 team. The project we will present at the Giant Jamboree relies on a plasmid that will give bacteria the ability to express specific proteins as a function of temperature. We aim to make the bacteria produce a particular protein only at low temperatures (~10-15°C) and another protein only at high temperatures (~38-42°C). Here we will describe only the molecular modeling of the regulatory RNA sequence responsible for cold induced expression.

The cold sensing mechanism relies on the formation of specific structures at low temperatures due to stabilization of specific types of RNA, more generally called RNA thermometers. The translation of cspA (cold-shock protein A of *E. coli*) is thought to be dependent on the conformation of the cspA 5'UTR. The main hypothesis is that at high temperature (around 37°C) the cspA 5'UTR is rapidly degraded, preventing cspA expression. At low temperature (below 15°C) the cspA 5'UTR is stable and differentially favors ribosome binding, likely through enhanced recognition of ribosomal protein S1 in the pre-initiation complex. Since RNA conformation and stability are difficult to study experimentally, we decided to model it computationally using the SimRNA software package[1].

We used SimRNA to generate a 3D model of the 5' UTR of the cspA mRNA to study its possible structure. Our work also led us to use several additional modeling software packages such as GROMACS [2]and VMD[3]. We wrote this handbook to help the iGEM community avoid the delays we faced early in our modeling work.

[1] Michal Boniecki and al., "SimRNA: A Coarse-Grained Method for RNA Folding Simulations and 3D Structure Prediction," accessed July 20, 2017, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4838351/.

[2] Mark James Abraham et al., "GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers," *SoftwareX* 1–2 (September 2015): 19–25, https://doi.org/10.1016/j.softx.2015.06.001.

[3] William Humphrey, Andrew Dalke, and Klaus Schulten, "VMD: Visual Molecular Dynamics," *Journal of Molecular Graphics* 14, no. 1 (February 1, 1996): 33–38, https://doi.org/10.1016/0263-7855(96)00018-5.

The following document will start by explaining the principle and use of the SimRNA software to do the RNA simulation. It will be followed by the conversion of the SimRNA results into files readable by VMD using GROMACS. Finally, we will talk about the use of VMD for visualization and analysis.

For additional information, please contact us at: ionis.igem@gmail.com or visit our wiki: http://2017.igem.org/Team:IONIS-PARIS .

Software installation and command line basics on Linux

To run our simulations, we used a personal gaming PC running Linux 16.04 LTS.
Specs:

CPU: Intel Core i7 6700K @ 4.00GHz
RAM: 16.0 GB Dual-Channel DDR4
Graphics card: NVIDIA GeForce GTX 1070 - 8GB

Motherboard: ASUSTeK COMPUTER INC. Z170 PRO GAMING (LGA1151)

VNC (Virtual Network Computing)

Virtual Network Computing is a type of software that will allow the access of your server via the internet from anywhere in the world. In our case, it allowed us to manage our simulations (some of which took several days to complete) remotely.

Today several VNC software implementations are available; some are free, but for convenience we used VNC connect from RealVNC with the free trial. They have easy to follow tutorials for installation and VNC configuration.

SimRNA

To download SimRNA for Linux go to: ftp://ftp.genesilico.pl/pub/software/simrna/. Choose the file that suits your machine and extract it.

It's done! (easy huh?)

VMD

You will find all the versions of VMD available on this website:
http://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=VMD

To download VMD you need to create a free VMD user account. Then extract the installation file.

GROMACS

Before setting up GROMACS, run those three prerequisite installations:

```
sudo apt-get install cmake
sudo apt-get install make
sudo apt-get install build-essential gcc
```

To install GROMACS, go on http://manual.gromacs.org/documentation/. Choose the package you want to install (for ou simulations we used the 2016.3 version) and extract the archive on the desktop.

Open a terminal and navigate to the new GROMACS folder with this line:

cd path_to_gromacs_folder.

Then make a new directory with *mkdir build* and enter in this new directory *cd build*.

Now, let's get serious:

CMake will allow to choose the options to add to your own GROMACS compiled version. The options will depend on your hardware and your software.

Some usefull options:

- If you are using a multi-threading processor you can build GROMACS with MPIsupport: -DGMX_MPI=on

- If you are using a NVIDIA graphics card with CUDA GPU acceleration you can add: -DGMX_GPU=on. We advise you to install the CUDA toolkit and add to your CMake command: -DCUDA_TOOLKIT_ROOT_DIR=path_to_cuda_toolkit_folder (usually /usr/local/cuda).

- To allow you to do the benchmark add to your CMake command: -DREGRESSIONTEST_DOWNLOAD=on and -DMGX_BUILD_OWN_FFTW=on.

For more options, we advise you to refer to the GROMACS manual in the section "Typical installation".

For our CMake command we have: cmake ..-DGMX_MPI=on -DCUDA_TOOLKIT_ROOT_DIR=/usr/local/cuda -DREGRESSIONTEST_DOWNLOAD=on -DMGX_BUILD_OWN_FFTW=on

Then run three commands:

make
make check
sudo make install

Then to launch GROMACS just run *gmx* in a terminal.

Simulations and rendering

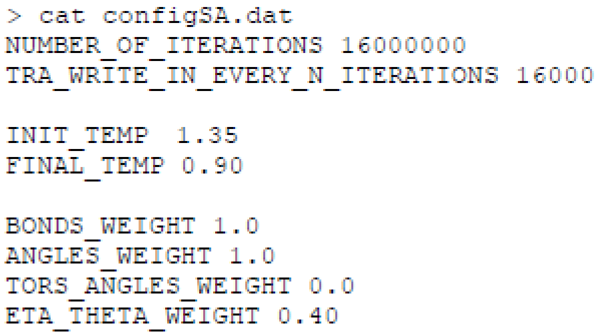
Run a simulation

Isotemperature simulations

To run a simulation, you can refer to the section 3 of the SimRNA User Manual and use the appropriate command. In our experiments we used a .txt file but you can also use a .seq or a PDB file.

Code for a .txt file: SimRNA -s input_file_seq -c config_file

-c : it gives a config file.



```
> cat configSA.dat
NUMBER_OF_ITERATIONS 16000000
TRA_WRITE_IN_EVERY_N_ITERATIONS 16000

INIT_TEMP  1.35
FINAL_TEMP 0.90

BONDS_WEIGHT 1.0
ANGLES_WEIGHT 1.0
TORS_ANGLES_WEIGHT 0.0
ETA_THETA_WEIGHT 0.40
```

And then wait… ⍰ (Depending on system size and hardware specifications, this could take up to several days for a large RNA, such as ours of ~190 nucleotides).

Once the simulation is finished you can have a quick view of the results with TraflView, SimRNA's built in trajectory viewer. Then if you want to get a .vmd movie of the RNA folding continue to Extract the .trafl and the .xtc .

Code: traflView output.trafl output.bonds

REMC (Replica Exchange Monte Carlo) simulations

For the REMC simulationsjust use the same previous code and add -E that allows to specify the number of replicas you want.

Don’t forget to modify the config file to choose your temperature range.

Once the simulation is finished, you have to do the clustering:

Begin by concatenate all the trafl. Go into the trafl directory (*cd trafl_folder*).

And run in a terminal: *cat *.trafl > output.trafl*

Go back into the SimRNA root folder:

```
clustering trajectory.trafl X.X X.X >& remc_clust.log
```

- *>& remc_clust.log* : allow to generate a log file

- *X*.X indicate the percentage of the lowest energy molecule that will be included into the clustering. If X.X = 0.01 it means that 1% of the lowest energy frames of the generated trafl will be included in the cluster.

Each trafl can be transformed into xtc by following the section Extract the .trafl and the .xtc. As the reference pdb for the trafl2pdbs command you can use the first pdb file generated by the simulation. No matter the temperature of the first pdb file, the simulation always begins with a circular RNA.

Extracting the trajectory files : .trafl and .xtc

After the simulation, the program output that interests us is the TRAFL, that is a trajectory file. We now have to extract the PDB files from the trafl. Every PDB file represent one frame of the simulation.

```
Code: ./SimRNA_trafl2pdbs input_file.pdb output_trafl_file.trafl 000:000 AA
```

- *input_file.pdb*: It corresponds to the PDB file that you gave to SimRNA to begin the simulation.

- *output_trafl_file.trafl*: It is the TRAFL fil that SimRNA generated at the end of the simulation

- *000:000*: Optional. It allows to choose the frames we want to extract into PDB files. It can be useful if your total number of frame is over 1020. Indeed, we didn’t succeed in extracting more than this number of frame in one time.

- *AA*: Optional. This allow to render a PDB file of each frame with a full atom reconstruction. All the file generated will be named as follow: _AA.pdb

Each frame, with the AA option, will generate 2 PDB files and 1 ss_detected file.

Then isolate all the _AA.pdb files in a directory with the command:

```
Code : cd ./your_pdb_folder
        mkdir AA_pdb
        mv *AA.pdb ./AA_pdb
        rm *.pdb & rm *.ss_detected
```

Those commands will create a new folder "AA_pdb", move all the AA pdb files in it, and clean the original folder from all the .pdb & .ss_detected files.

Generate the .xtc file, another trajectory file, which is supported by VMD.

You need to use GROMACS and work in the directory where the _AA.pdb files are stored.

```
Code : gmx trjcat -f *AA.pdb -o file_name.xtc -cat
```

- *-f *.pdb* : the -f is used to give the name of the input file. The *AA.pdb allow to concatenate all the _AA.pdb files into the .xtc file.

- *-o file_name.xtc*: the -o is used to give the name of the output file.

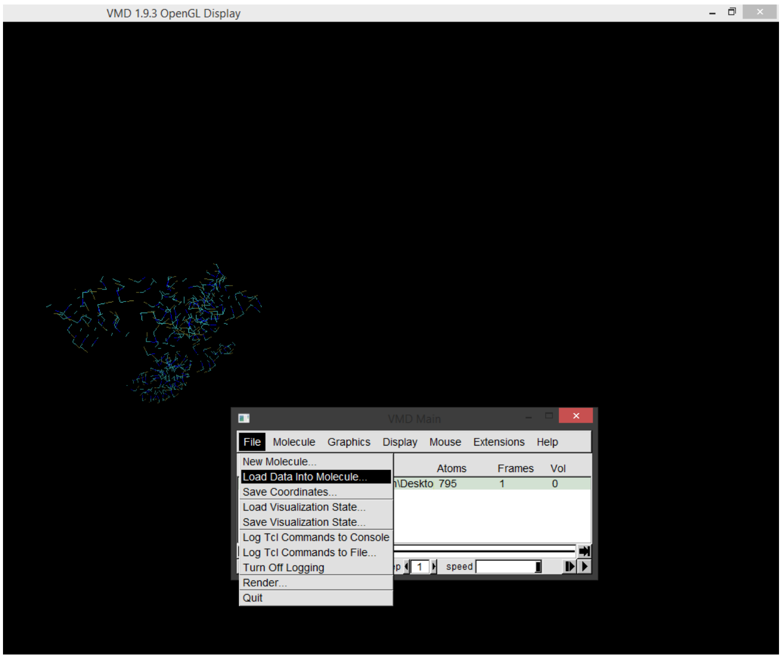
- -cat : allow to concatenate all the frames.

Visualization on VMD

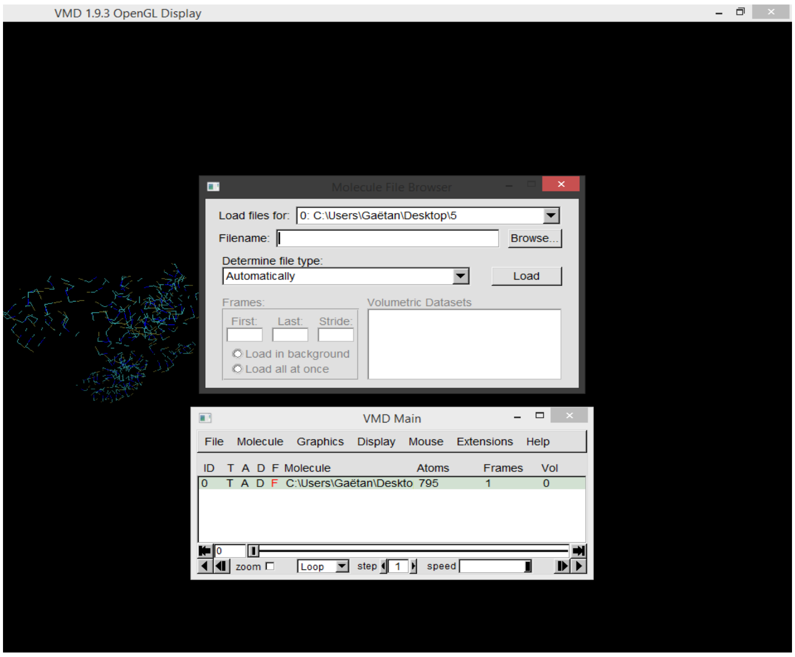
The visualization part can be done either on windows or Linux.

Tips : To launch VMD on linux just write *vmd* in a terminal.

Open VMD and load the first AA pdb file of the simulation. Then load the xtc file. It could take a while but to accelerate the process you can hide the molecule. For that you double click on the name of the molecule:



File > Load Data into Molecule…


Load files for: 0: C:\Users\Gaëtan\Desktop\5

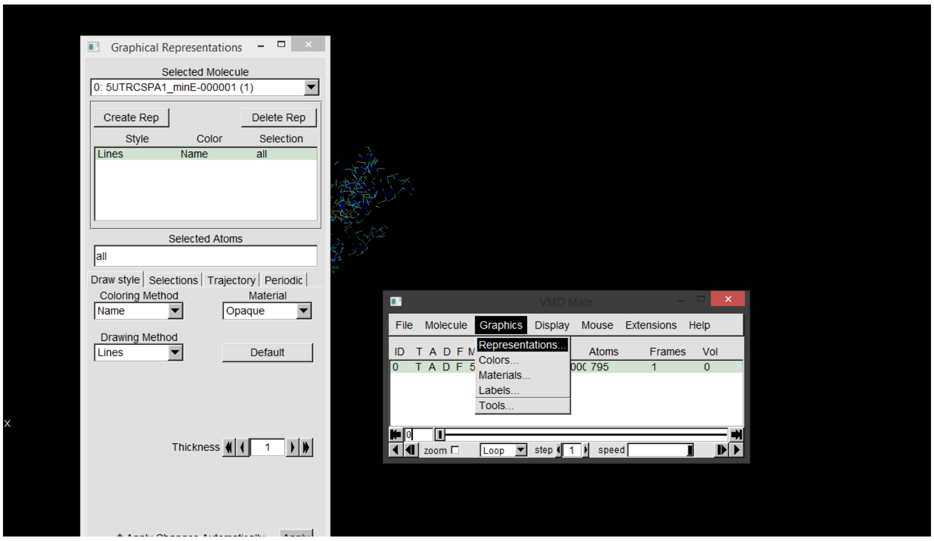
Filename:

Determine file type: Automatically

ID	T	A	D	F	Molecule	Atoms	Frames	Vol
0	T	A	D	F	C:\Users\Gaëtan\Deskto	795	1	0

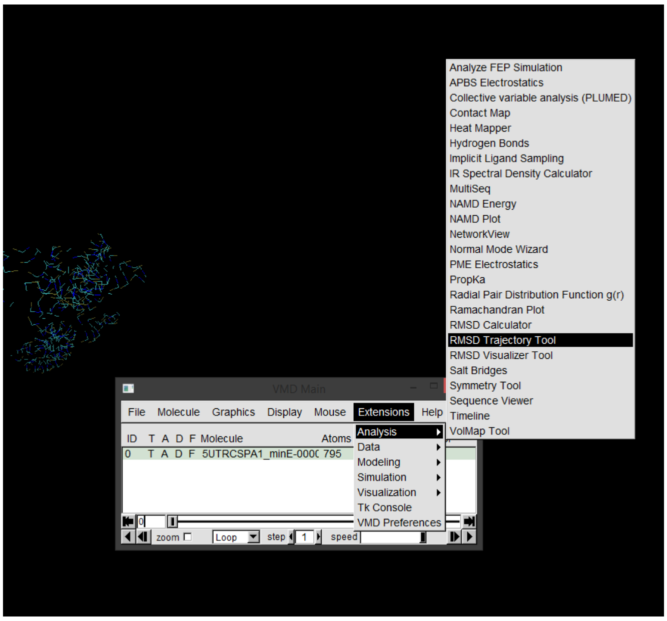
Verify you *Load files for* your molecule and *Browse* the xtc.

VMD will load all the frames of your molecule, if you have a large number of frame you can hide the molecule in order to accelerate the loading process:

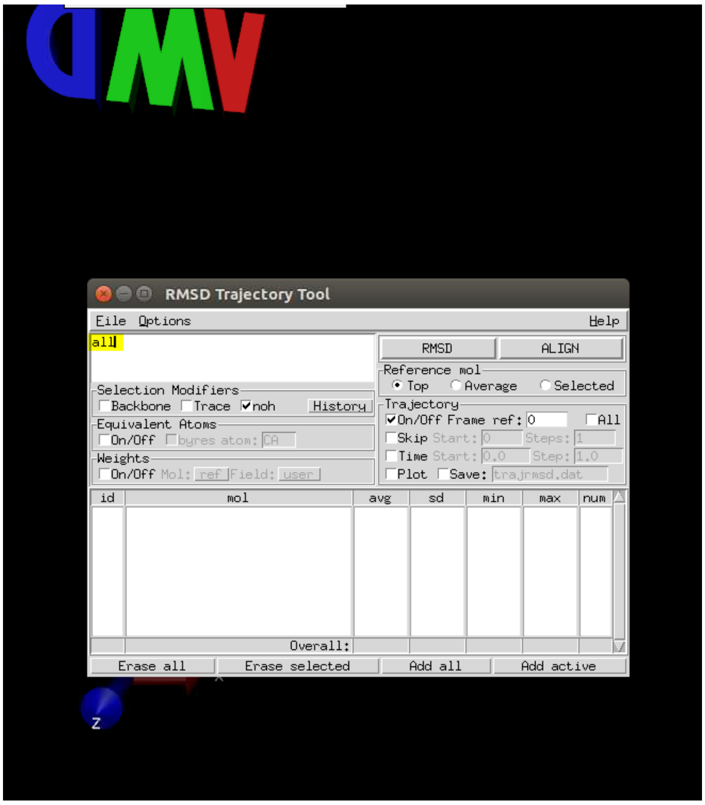


Graphics > Representation > Delete Rep. At the end of the loading process you can follow the same path and chose *Create Rep* in order to visualize the molecule.

Center the molecule:



Extensions > Analysis > RMSD Trajectory Tool


VMD
RMSD Trajectory Tool
File Options Help
all
RMSD
ALIGN
Reference mol
● Top ○ Average ○ Selected
Selection Modifiers
□ Backbone □ Trace ☑ noh History
Trajectory
☑ On/Off Frame ref: 0 □ All
Equivalent Atoms
□ On/Off □ byres atom: CA
□ Skip Start: 0 Steps: 1
□ Time Start: 0.0 Step: 1.0
Weights
□ On/Off Mol: ref Field: user
□ Plot □ Save: trajrmsd.dat
id mol avg sd min max num
Overall:
Erase all
Erase selected
Add all
Add active
z

Write “all” and press Align.

Then to reset the view and put the molecule in the center press “=” on your keyboard or go to *Display > Reset view*.

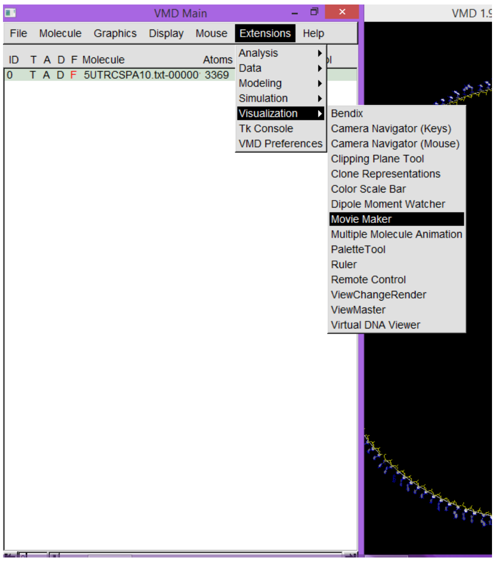
Video making on Windows

First, you must download the VideoMach software. It is used by VMD to render the movie.

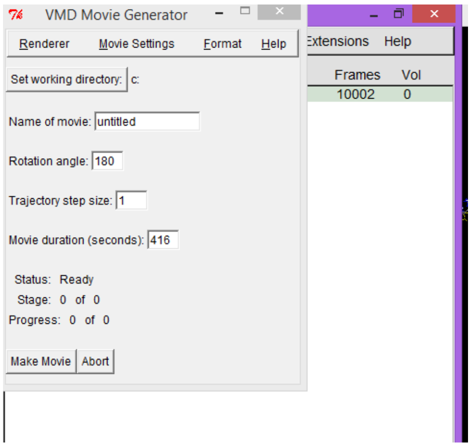
Unlike what VMD tells you, the free version works just fine, no need to pay $20 per month.

Once it is done. Load your PDB and your XTC as previously described.

Now, Select Extensions → Visualization → Movie Maker.



Set the output directory and the movie name.



Set working directory: c:

Name of movie: untitled

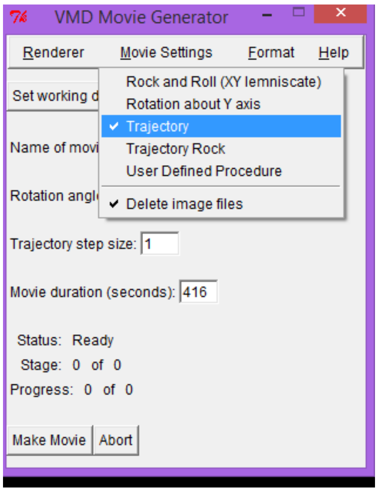
Rotation angle: 180

Trajectory step size: 1

Movie duration (seconds): 416

Status: Ready
Stage: 0 of 0
Progress: 0 of 0

In the menu Movie Settings, select the option Trajectory, you cannot edit the movie duration when you are recording a trajectory.



Trajectory step size: 1

Movie duration (seconds): 416

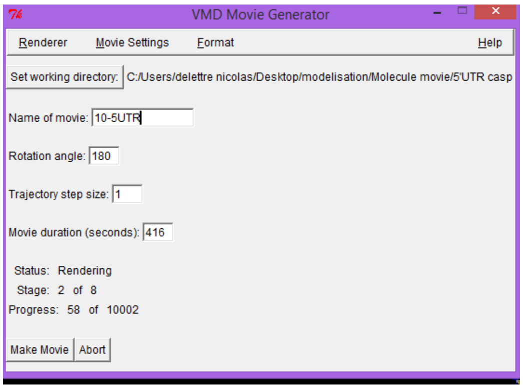
Status: Ready

Stage: 0 of 0

Progress: 0 of 0

Press “MAKE A MOVIE”.

Wait until step 2/8 is complete (it renders all your frames into screenshots).



Set working directory: C:/Users/delettre nicolas/Desktop/modelisation/Molecule movie/5'UTR casp

Name of movie: 10-5UTR

Rotation angle: 180

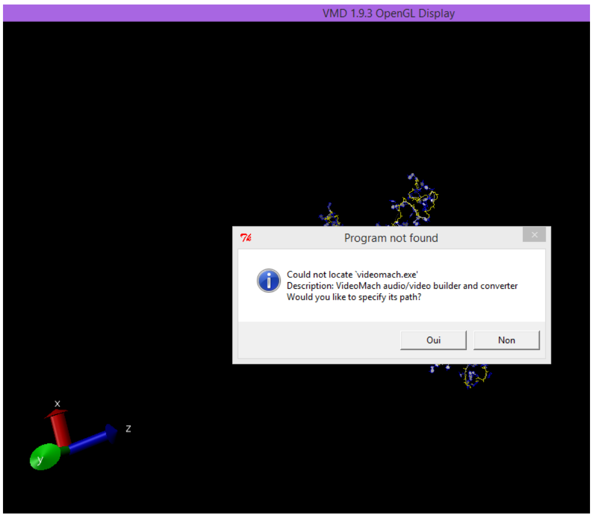
Trajectory step size: 1

Movie duration (seconds): 416

Status: Rendering

Stage: 2 of 8

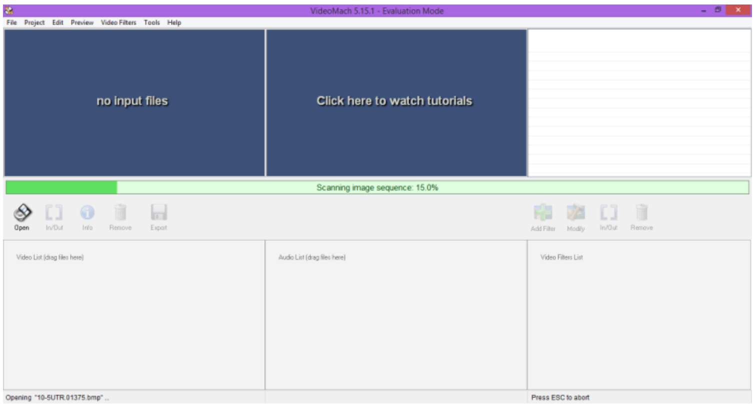
Progress: 58 of 10002



Could not locate `videomach.exe'
Description: VideoMach audio/video builder and converter
Would you like to specify its path?

When asked, indicate the pathway towards VideoMach executable (videomach.exe).

It is probably “C:\Program Files (x86)\VideoMach\videomach.exe”.

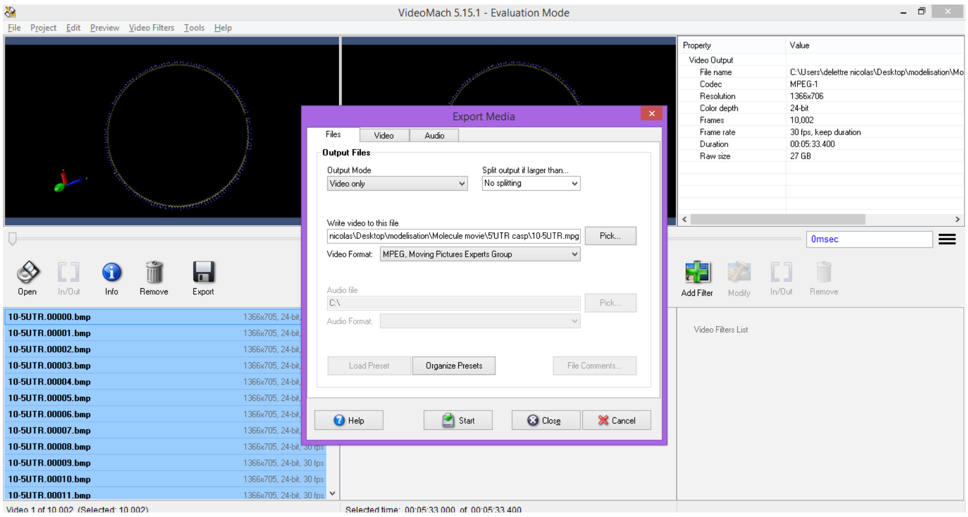

VideoMach 5.15.1 - Evaluation Mode
File Project Edit Preview Video Filters Tools Help
no input files
Click here to watch tutorials
Open
In/Out
Info
Remove
Export
Add Filter
Modify
In/Out
Remove
Video List (drag files here)
Audio List (drag files here)
Video Filters List

Scanning image sequence: 15.0%

Opening "10-5UTR.01375.bmp" ...

Press ESC to abort

VideoMach will analyze the image sequence, just wait a bit.



Property	Value
Video Output	
File name	C:\Users\delettre nicolas\Desktop\modelisation\Mo
Codec	MPEG-1
Resolution	1366x706
Color depth	24-bit
Frames	10,002
Frame rate	30 fps, keep duration
Duration	00:05:33.400
Raw size	27 GB

0msec

Output Files

Output Mode: Video only

Split output if larger than...: No splitting

Write video to this file: nicolas\Desktop\modelisation\Molecule movie\5'UTR casp\10-5UTR.mpg

Video Format: MPEG, Moving Pictures Experts Group

Audio file: C:\

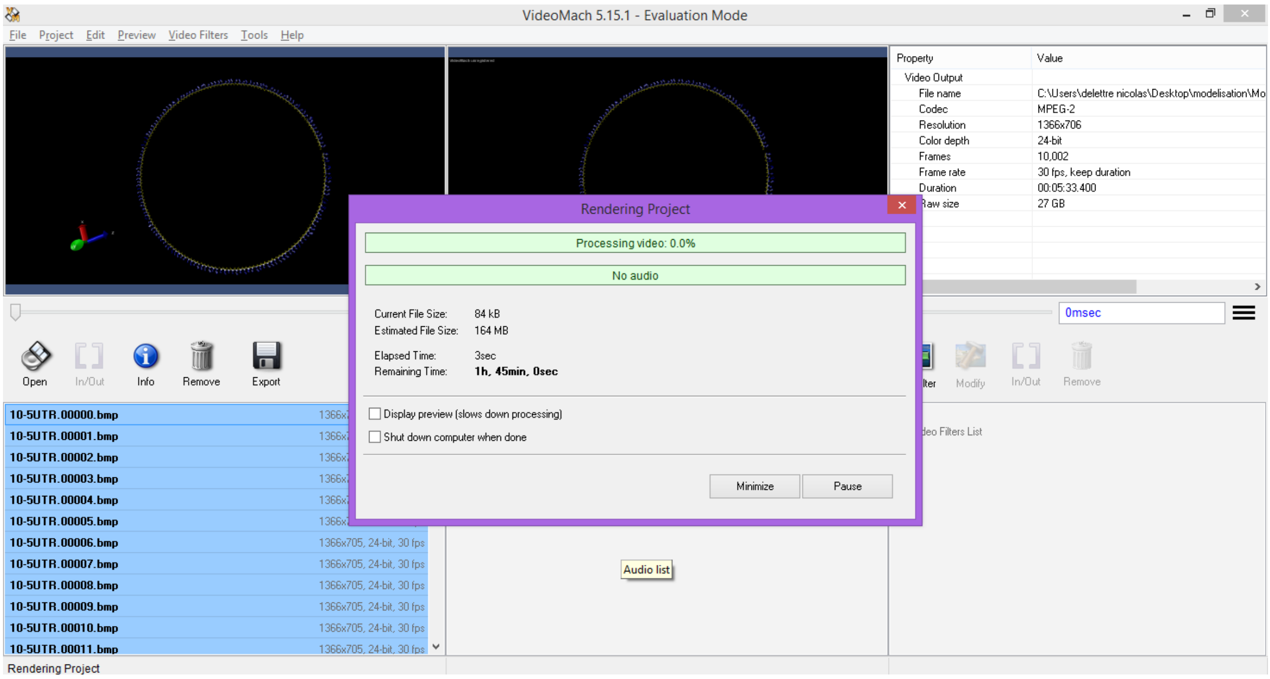
Audio Format:

10-5UTR.00000.bmp	1366x705, 24-bit,
10-5UTR.00001.bmp	1366x705, 24-bit,
10-5UTR.00002.bmp	1366x705, 24-bit,
10-5UTR.00003.bmp	1366x705, 24-bit,
10-5UTR.00004.bmp	1366x705, 24-bit,
10-5UTR.00005.bmp	1366x705, 24-bit,
10-5UTR.00006.bmp	1366x705, 24-bit,
10-5UTR.00007.bmp	1366x705, 24-bit,
10-5UTR.00008.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00009.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00010.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00011.bmp	1366x705, 24-bit, 30 fps

Video 1 of 10,002 (Selected: 10,002)

Selected time: 00:05:33.000 of 00:05:33.400

Just hit the start button to launch the final rendering.



Property	Value
Video Output	
File name	C:\Users\delettre nicolas\Desktop\modelisation\Mo
Codec	MPEG-2
Resolution	1366x706
Color depth	24-bit
Frames	10,002
Frame rate	30 fps, keep duration
Duration	00:05:33.400
Raw size	27 GB

Rendering Project

Processing video: 0.0%

No audio

Current File Size: 84 kB
Estimated File Size: 164 MB

Elapsed Time: 3sec
Remaining Time: 1h, 45min, 0sec

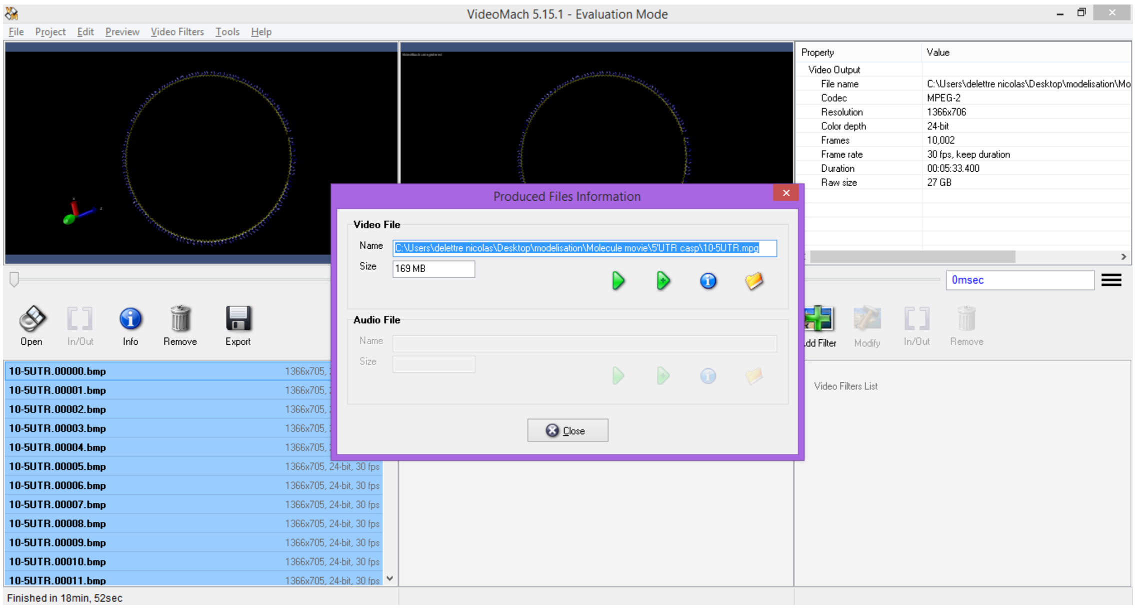
☐ Display preview (slows down processing)
☐ Shut down computer when done

Minimize Pause

10-5UTR.00000.bmp	1366x
10-5UTR.00001.bmp	1366x
10-5UTR.00002.bmp	1366x
10-5UTR.00003.bmp	1366x
10-5UTR.00004.bmp	1366x
10-5UTR.00005.bmp	1366x
10-5UTR.00006.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00007.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00008.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00009.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00010.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00011.bmp	1366x705, 24-bit, 30 fps

Rendering Project

Just wait until the end of the process.



Property	Value
Video Output	
File name	C:\Users\delettre nicolas\Desktop\modelisation\Mo
Codec	MPEG-2
Resolution	1366x706
Color depth	24-bit
Frames	10,002
Frame rate	30 fps, keep duration
Duration	00:05:33.400
Raw size	27 GB

Produced Files Information

Video File
Name C:\Users\delettre nicolas\Desktop\modelisation\Molecule movie\5'UTR casp\10-5UTR.mpg
Size 169 MB

Audio File
Name
Size

Close

10-5UTR.00000.bmp	1366x705,
10-5UTR.00001.bmp	1366x705,
10-5UTR.00002.bmp	1366x705,
10-5UTR.00003.bmp	1366x705,
10-5UTR.00004.bmp	1366x705,
10-5UTR.00005.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00006.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00007.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00008.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00009.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00010.bmp	1366x705, 24-bit, 30 fps
10-5UTR.00011.bmp	1366x705, 24-bit, 30 fps

Finished in 18min, 52sec

You should get your movie at the selected directory.

Bibliography

Abraham, Mark James, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. “GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers.” *SoftwareX* 1–2 (September 2015): 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

Boniecki, Michal, and al. “SimRNA: A Coarse-Grained Method for RNA Folding Simulations and 3D Structure Prediction.” Accessed July 20, 2017. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4838351/.

Humphrey, William, Andrew Dalke, and Klaus Schulten. “VMD: Visual Molecular Dynamics.” *Journal of Molecular Graphics* 14, no. 1 (February 1, 1996): 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.