**APPENDIX**

**Tutorial for CABS-dock results visualization and analysis**

Table Of Contents

# 1   Introduction

This tutorial describes how to:
- load all the necessary files into the VMD molecular graphics program
- create simple graphic representations for viewing trajectories
- align models from trajectories with a known receptor structure
- calculate peptide RMSDs (root mean square deviation from the reference structure)
- create plots from calculated RMSDs vs. energies from the CABS-dock results

The following software is required:
- VMD with an RMSD Trajectory Tool plug-in (included in VMD ver. 1.8.8 and higher). VMD is user friendly yet advanced software for the analysis and visualization of structure and molecular dynamics trajectories of biological systems. In this tutorial some features are omitted, e.g. VMD for basic protein analysis: Sequence Viewer, Contact Map or Ramachandran Plot. Those plug-ins are very well documented on the VMD website.
- gnuplot· gnuplot is software for creating high quality plots from user data.

Both programs are free to use, available for major operating systems and fairly easy to install.

## 2   CABS-dock result files and reference structure

In this tutorial, we used an example prediction run described in the manuscript (in section 3.1.1). The example was created using the following input data:

- **Peptide sequence:** SSRFESLFAG
- **Peptide secondary structure:** CHHHHHHHHC
- **Receptor input structure:** PDB ID: 2AM9, crystal structure of human androgen receptor in the unbound form (without a peptide)

Reference structure used for the calculation of RMSD values to the experimentally determined peptide-bound structure:

- **Peptide-receptor complex structure:** PDB ID: 1T7R, crystal structure of human androgen receptor in complex with the peptide

Typically the structure of the bound complex is unknown, so one of the resulting models may be used as a reference (for example the top scored model: model_1.pdb).

The results page of this job is available at:
http://biocomp.chem.uw.edu.pl/CABSdock/job/7f0bda72050182.

First, from the "Docking predictions results" tab, download a compressed archive with the output data for further analysis. The archive contains the following files:

- `model_*.pdb` – 10 final models, numbered from 1 to 10 (in the PDB file format and all-atom representation)
- `cluster_*.pdb` – cluster models (groups of models that have been classified in structural clustering to particular clusters), numbered from 1 to 10 (in the PDB file format and C-alpha representation). Cluster numbering corresponds to models numbering, i.e. model_7.pdb is a representative model for models grouped in the seventh cluster (ranked as seventh) (cluster_7.pdb).
- `trajectory_*.pdb` – complete set of 10 trajectories, numbered from 1 to 10 (in the PDB file format and C-alpha representation). Each trajectory contains 1000 models.
- `top1000.pdb` – top 1000 models (selected for further clustering and analysis) from the 10 trajectories (in the PDB file format and C-alpha representation)
- `input.pdb` – input structure of the receptor
- `README` – log file with all information to recreate the simulation
- `energy.txt` – log file with energy values (from the coarse-grained CABS force-field) for all resulting models; columns contain the following data (in the order listed):
  - trajectory (replica) index (`no`)
  - trajectory (replica) frame index (`fr`)
  - temperature - a parameter in the CABS model that controls the acceptance of new conformations within the Monte Carlo method (`temp`)
  - energy value for the protein receptor only (`recE`)
  - energy value for the peptide only (`pepE`)
  - energy value for the protein-peptide interaction only (`rec-pepE`)
  - energy value for the entire protein-peptide complex (`totalE`)

The energy.txt file contains the following data (column headers are described above):

```
no    fr   temp    recE      pepE    rec-pepE   totalE
 1     1   1.95 -2261.77  -37.87      0.00  -2299.64
 1     2   1.95 -2255.36  -38.43      0.00  -2293.79
 1     3   2.45 -2195.35  -21.44      0.00  -2216.79
 1     4   2.45 -2193.94    0.49   - 17.80  -2211.25
 1     5   2.95 -2160.39  -36.25     -5.60  -2202.24
                           ...
10   996   4.00 -1920.98  -14.86    -32.10  -1967.94
10   997   4.50 -1965.89   -3.03    -30.00  -1998.92
10   998   4.00 -1959.18   -5.52    -42.40  -2007.10
10   999   4.00 -1995.16  -14.23    -32.20  -2041.59
10  1000   4.00 -1961.22  -14.34    -43.60  -2019.16
```

Finally, download the 1T7R structure from the RCSB PDB database.

# 3   Using VMD

## Below, different font styles are used to mark different features:

- *Italics: buttons, menu positions, fields, lists, sections*
- `Lucida Console font: CONSOLE INPUT, VALUES, FILES`
- Underlined: VMD window panels

## 3.1   Loading files

Start VMD. Three application windows should be displayed: console, OpenGL Display and VMD Main. Load necessary files using one of the following methods: using the graphical interface (see section 3.1.1) or via the console (see section 3.1.2).

### 3.1.1   From the graphical interface

Click *File>New Molecule...*. From the Molecule File Browser window > select *Browse*. In the Choose a molecule file window navigate to server results, select file `trajectory_1.pdb` and click *Open* (or *OK*).
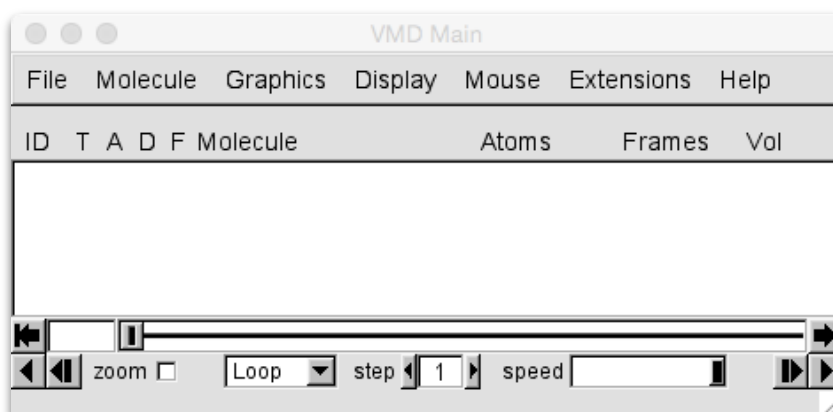
In the Molecule File Browser in the *Determine file type* field select PDB. Choose *Load all at once* in the *Frames* section and click *Load*.

To load the rest of trajectories select `trajectory_1.pdb` in the VMD Main window and choose *File>Load Data Into Molecule* and repeat previous steps in the Molecular File Browser window.

**It is important to load trajectories in the following order:** `trajectory_2.pdb`, `trajectory_3.pdb, ..., trajectory_10.pdb` **for further analysis (plotting RMSDs vs. Energy).**

The VMD Main window frame indicator should show: 9999 due to VMD numbering from 0 to 9999 for the total of 10000 frames. Load the reference structure (`1T7R.pdb`) as a new molecule. Set trajectories "top"- in the VMD Main window click under column marked "T" in the trajectories line. The "top" state is marked with "T"
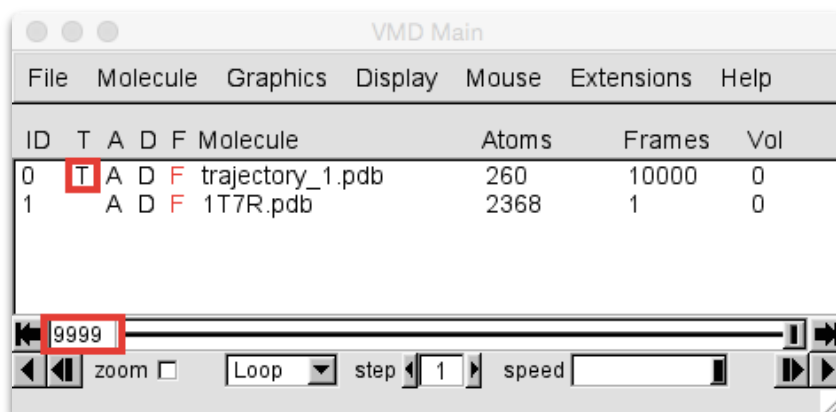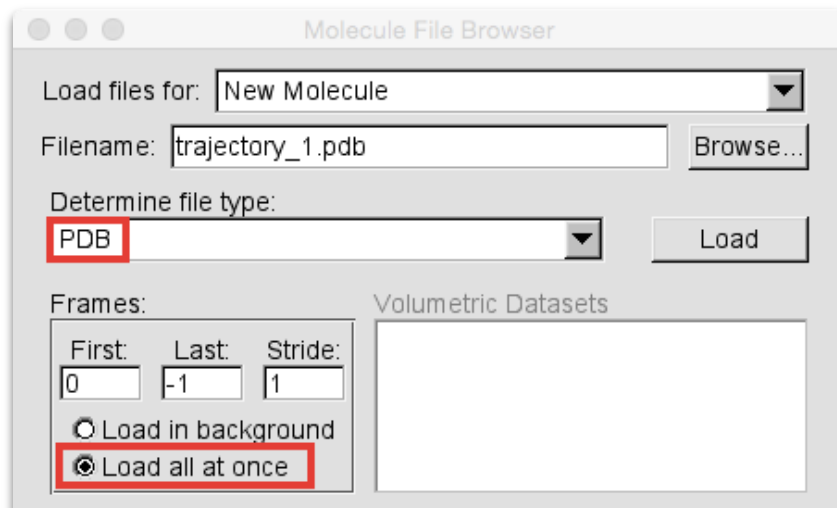
### 3.1.2  Via the console

An alternative (and faster) way of loading all files is through *Tk Console*.

From the VMD Main window click *Extensions>Tk Console*. In the VMD TkConsole window navigate, using the `cd` command (change directory – e.g.
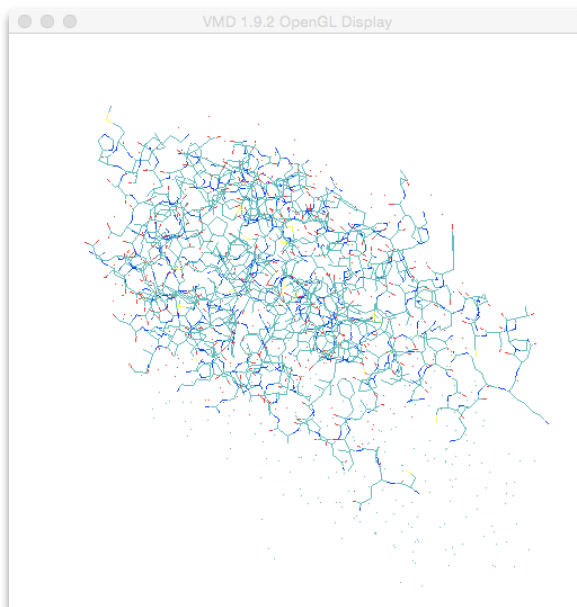`cd /home/user/Documents/results/`) to a folder with server results.
Then execute following commands:

```
for { set i 1 } { $i <= 10 } { incr i } {mol addfile trajectory_$i.pdb wait for
all }
mol new 1T7R.pdb
mol top
```
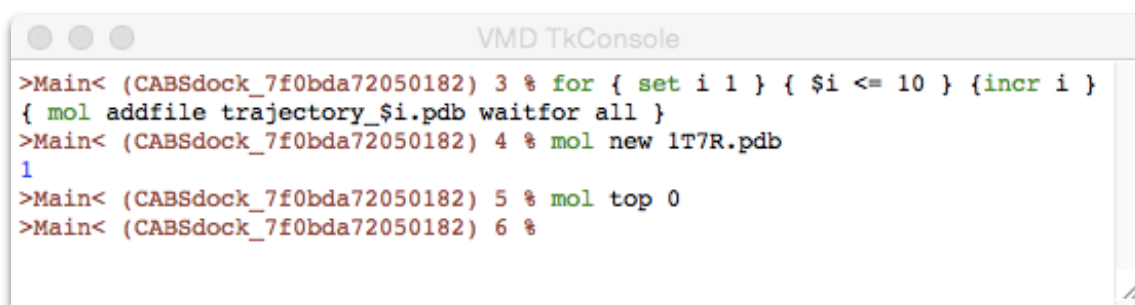
## 3.2 Changing graphical representations

This section describes how to change the default representation (see screenshot on the right) to a more convenient one.

### 3.2.1 Chain identification

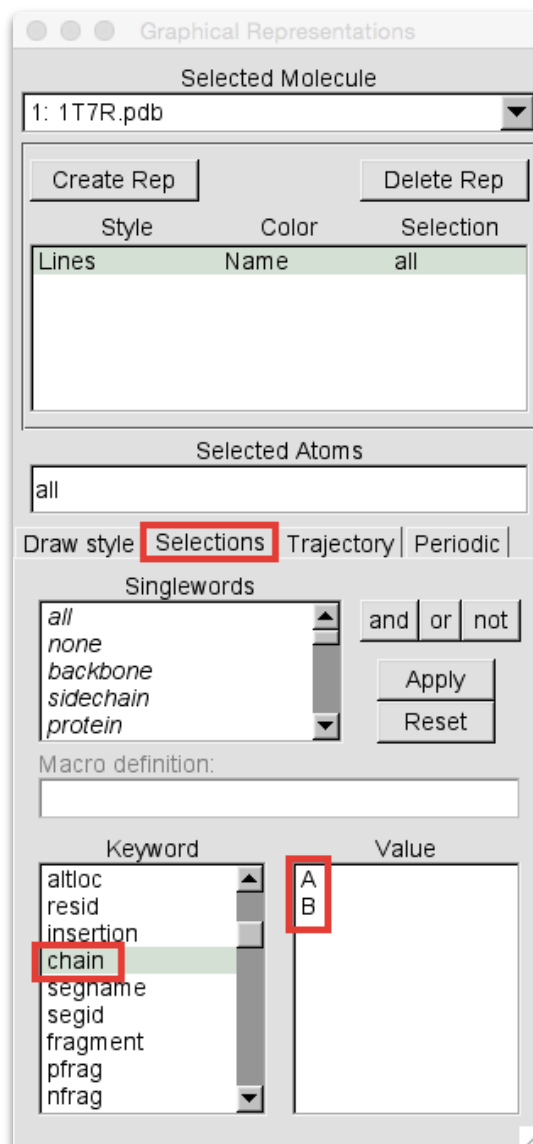First, it is necessary to find chain identifiers.

From the VMD Main window *Graphics*>select *Representations….*
In the Graphical Representations window choose *Selections* tab >*Keyword* list >*chain*.

The characters that appear in the Values section are chain IDs of the molecules in the currently selected molecule (*Selected Molecule* list).

**It is crucial to remember or note chain IDs for further analysis (molecule fragment selection).**

Find the chain IDs for the second molecule. In the example, both the trajectories and the reference contain chains A and B. A more complicated system may contain more chains in the receptor molecule but the peptide always has only one (usually the shortest) chain.

### 3.2.2 Creating trace representation

Now delete the existing representations by selecting entries on the list and click *Delete Rep*. In the *Selected Atoms* field select chain A and click *Create Rep*. For every representation set *Drawing Method* on *Trace*, *Coloring Method* on *ColorID* and choose color. Repeat the procedure for the rest of the chains and all other molecules from the *Selected Molecule* list. Now it is possible to view all the models from trajectories in the OpenGL Display window using the VMD Main frame slider.

Colors in the example screenshot below: red - 1T7R receptor, blue - receptor from the trajectory, green - 1T7R peptide, orange – peptide from the trajectory.
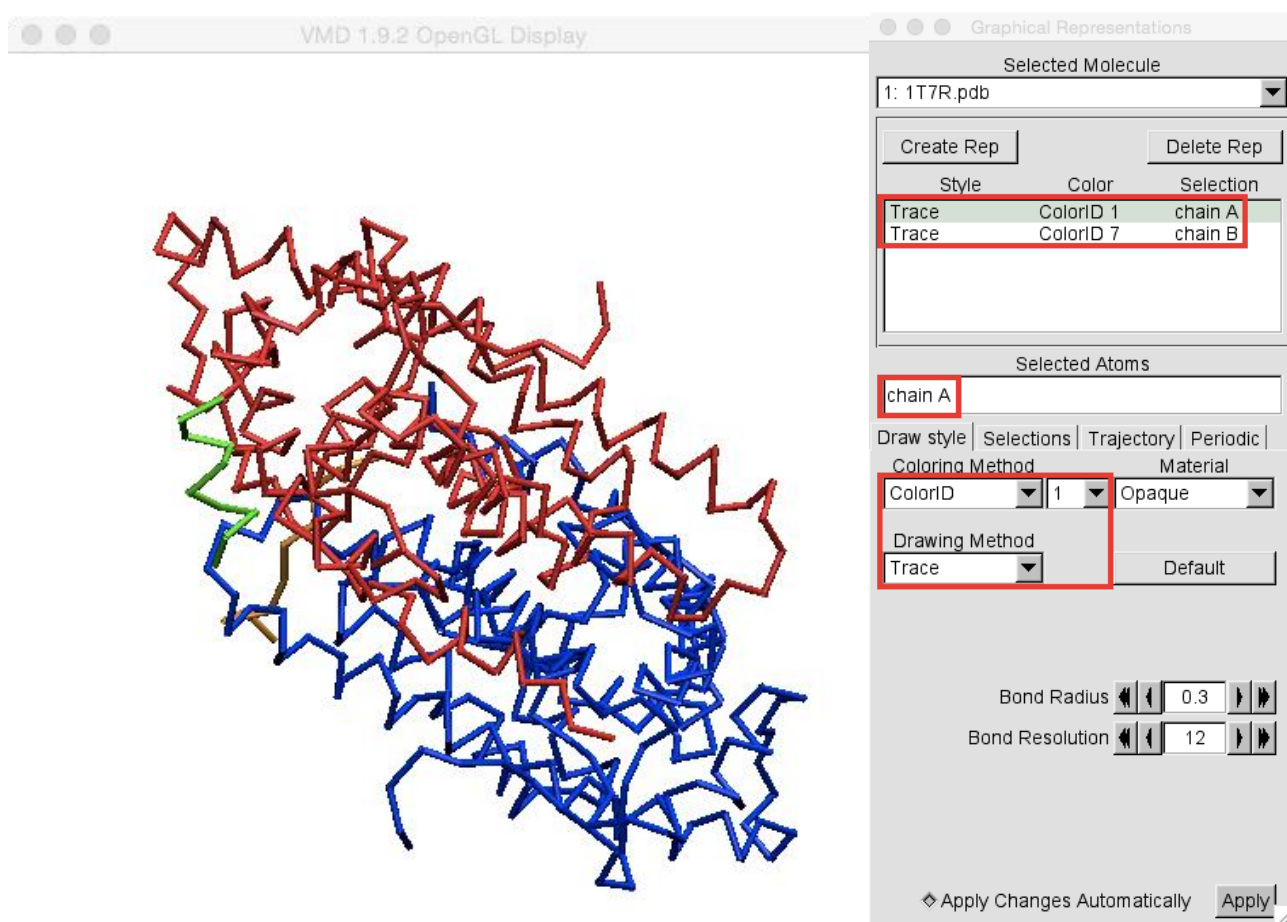
## 3.3  RMSD analysis

**This part concerns calculating peptide-based RMSDs through 10,000 models of an example trajectory.**
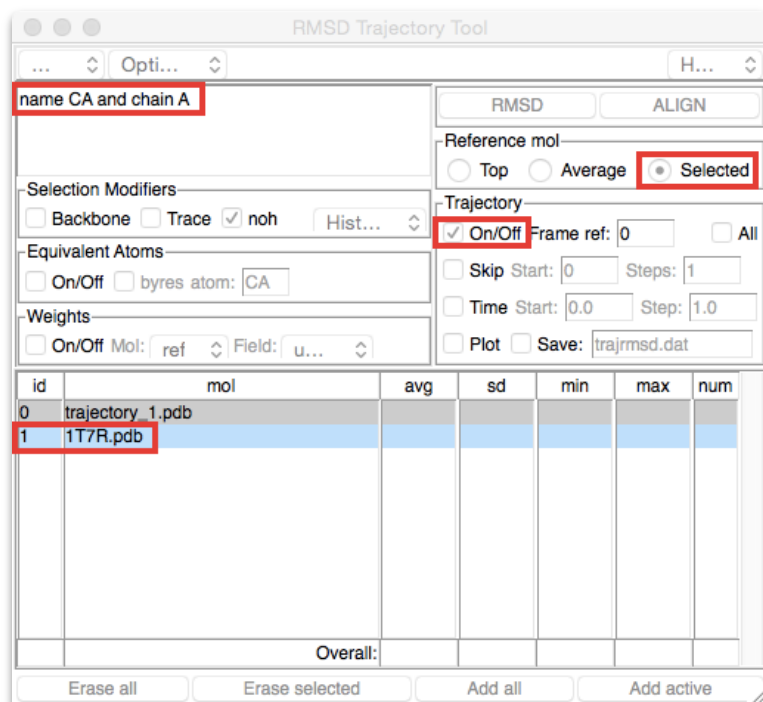


From the <u>VMD Main</u> window select *Extensions>Analysis>RMSD Trajectory Tool*. The <u>RMSD Trajectory Tool</u> window should contain names of two files:`trajectory_1.pdb` and `1T7R.pdb`. If not, click *Erase all* and *Add all*. In the *selection frame* (in the top left corner of the window) type: `name CA and chain A`. This command selects alpha carbon atoms from the protein backbone from chain A (receptor).

Set *Reference mol* to *Selected* and select the *Trajectory>On/Off* checkbox and click *Align*.
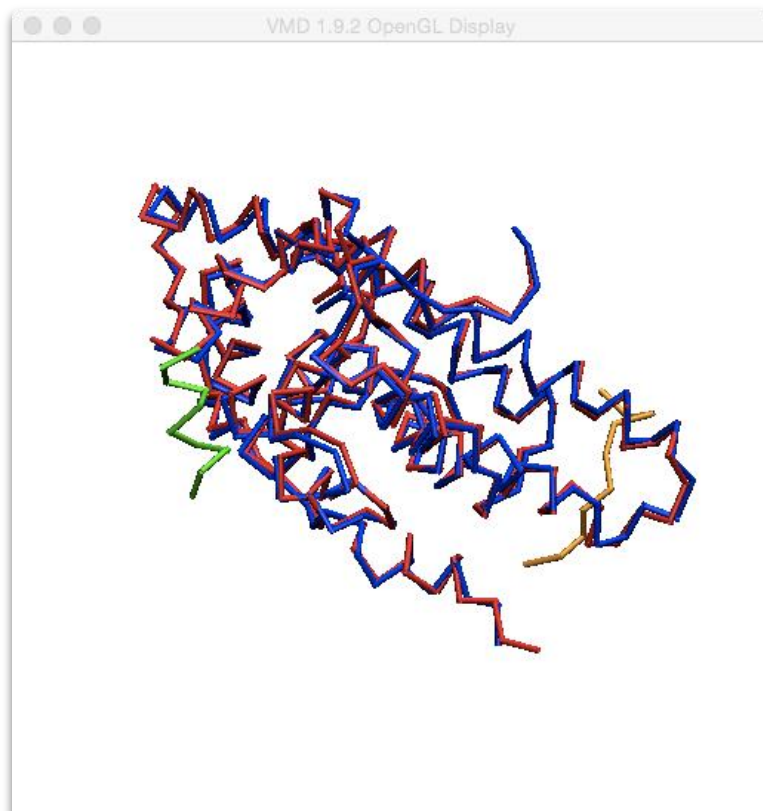


In the <u>OpenGL Display</u> window you will see that receptors from both systems are aligned.

Back in the <u>RMSD Trajectory Tool</u> window in the *selection frame* type: `name CA and chain B`. This selects alpha carbon atoms from the protein backbone from chain B (peptide). Click *RMSD.*

Finally, from the <u>RMSD Trajectory Tool</u> window menu choose *Save data....* and save data as a text file: `trajrmsd.txt` in the results folder.

The file contains frame IDs and calculated RMSDs:

```
frame    mol0
    0   34.489
    1   33.945
    2   25.595
    3   27.096
    4   30.941
      ...
 9997   48.781
 9998   45.021
 9999   42.475
```

It is also possible to view and save a simple RMSD vs. Frame plot by choosing *Plot data*.

## 3.4   Trajectory movie

After aligning receptor structures, it is easy to observe conformational sampling of the peptide. In the <u>OpenGL Display</u> window click and hold the left mouse button to rotate representations to expose the peptide from the reference structure.

Buttons in the <u>VMD Main</u> window: *Play forward* and *Play in reverse* allow the user to view the whole trajectory as a movie. To adjust speed or frame skipping use the speed slider and step counter.

To find a particular frame use *Step forward, Step in reverse* or use the frame slider by clicking on it and holding the left mouse button to drag the frame marker.
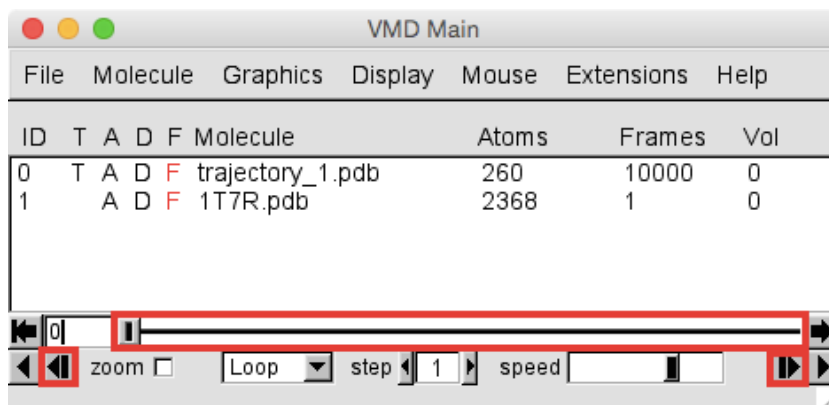


VMD also enables exporting the trajectory via the <span style="color:blue"><u>Movie Maker</u></span> plugin: <u>VMD Main</u>>*Extensions*>*Movie Maker* but it requires an external encoder (e.g. ffmpeg or mencoder for Linux/OS X, VideoMach for Windows) and preferably an external renderer.

# 4   Creating plots

## 4.1   File preparation

This part is about file preparation for gnuplot, for the Linux/OS X console. Windows users are encouraged to use text editors with block editing capabilities for merging files (e.g. Komodo Edit, Sublime Text 2 or Edit Plus, vim).

Before plotting with gnuplot the data from `energy.txt` and `trajrmsd.txt` have to be combined into one file. First remove the header from `trajrmsd.txt`:

```
sed 1d trajrmsd.txt > tmprmsd.txt
paste energy.txt tmprmsd.txt > enermsd.txt
```

The new file will resemble the sample (in the actual file there will be no header):

```
no    fr temp      recE     pepE rec-pepE    totalE    VMD-fr     RMSD
1     1 1.95 -2261.77 -37.87     0.00 -2299.64         0  46.085
1     2 1.95 -2255.36 -38.43     0.00 -2293.79         1  43.458
1     3 2.45 -2195.35 -21.44     0.00 -2216.79         2  43.565
1     4 2.45 -2193.94   0.49   -17.80 -2211.25         3  33.113
1     5 2.95 -2160.39 -36.25    -5.60 -2202.24         4  38.857
                                ...
10  998 4.00 -1959.18  -5.52   -42.40 -2007.10      9997  22.368
10  999 4.00 -1995.16 -14.23   -32.20 -2041.59      9998  17.720
10 1000 4.00 -1961.22 -14.34   -43.60 -2019.16      9999  16.552
```

The extreme values for RMSD  and energies can be found quickly using `sort` and `head` commands:

`sort -n -k9,9 enermsd.txt | head -n10` gives top 10 records with the lowest RMSD
`sort -n -k6,6 enermsd.txt | head -n10` gives top 10 records with the lowest interaction energy

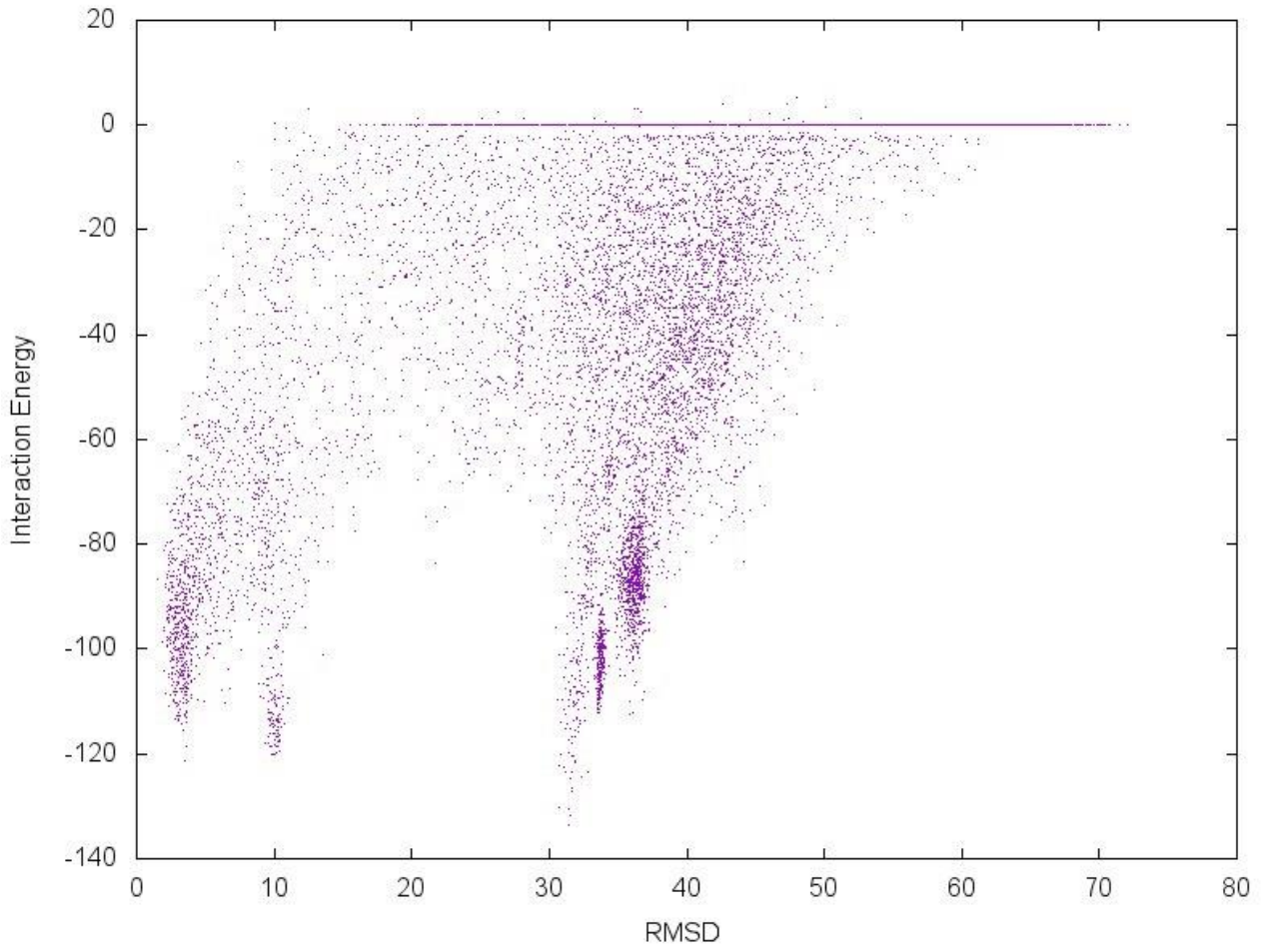The value in the `-k` parameter indicates which column is used for sorting.

## 4.2   Using gnuplot

Start gnuplot (e.g. by typing `gnuplot` in the console in the results folder) and set a proper terminal size, e.g. jpeg with 800x640 plot resolution:
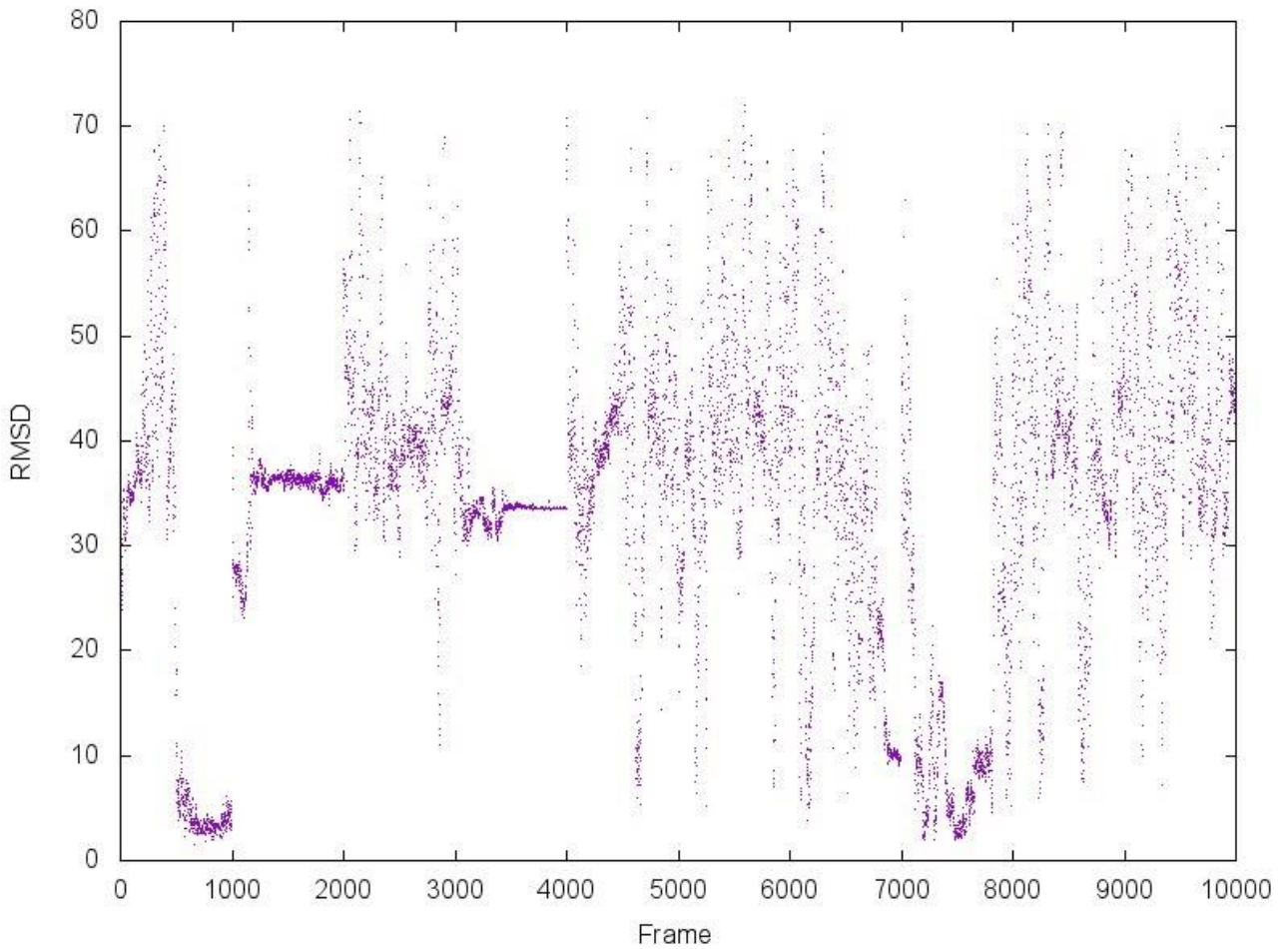
```
gnuplot> set terminal jpeg size 800,640
gnuplot> set output "rmsdene.jpg"
gnuplot> set xlabel "RMSD"
gnuplot> set ylabel "Interaction Energy"
gnuplot> set title "Energy vs RMSD"
gnuplot> plot "enermsd.txt" using 9:6 with dots notitle
```

This will produce a 800 by 640 JPEG `rmsdene.jpg` file with a dotted scatter plot titled "Energy vs. RMSD", its X axis labeled "RMSD" and Y axis labeled "Interaction Energy". Numerical values in the last command indicate the columns used in format "X:Y", so changing it to `8:9` will create a RMSD vs. frame plot and to `8:6` will create an Interaction energy vs. frame plot. Do not forget to also change the appropriate filename, title and labels. For other options please check the gnuplot [documentation.](#)

## Energy vs RMSD



## RMSD vs Frame

Energy vs Frame