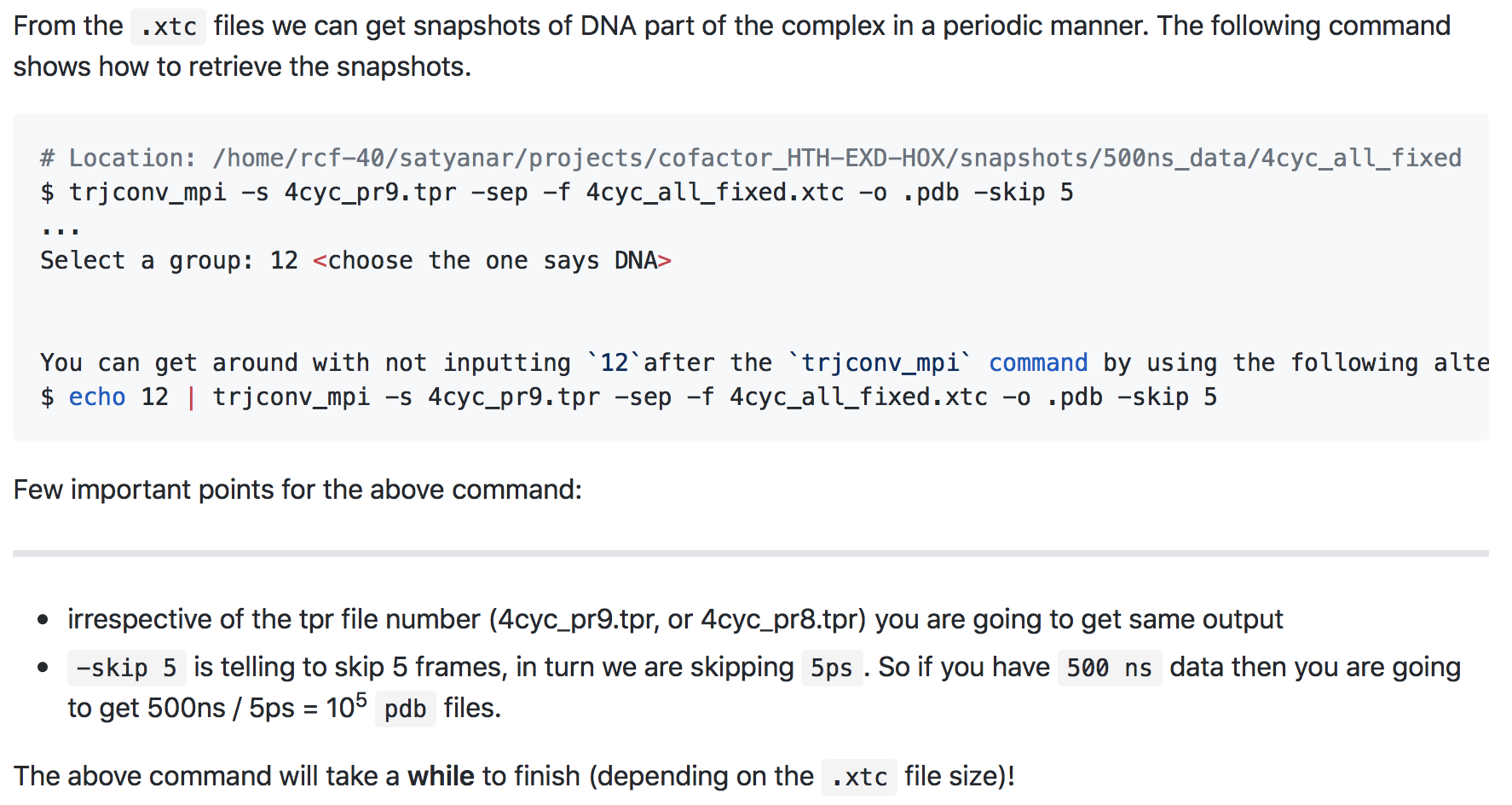
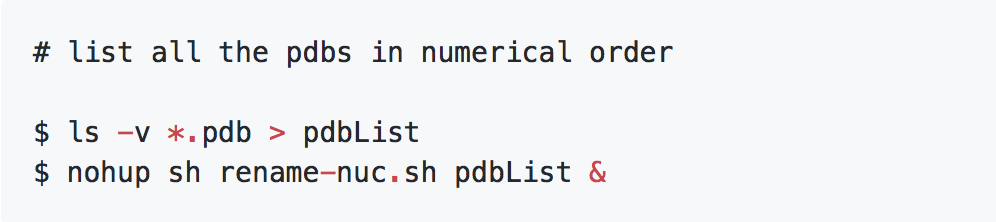
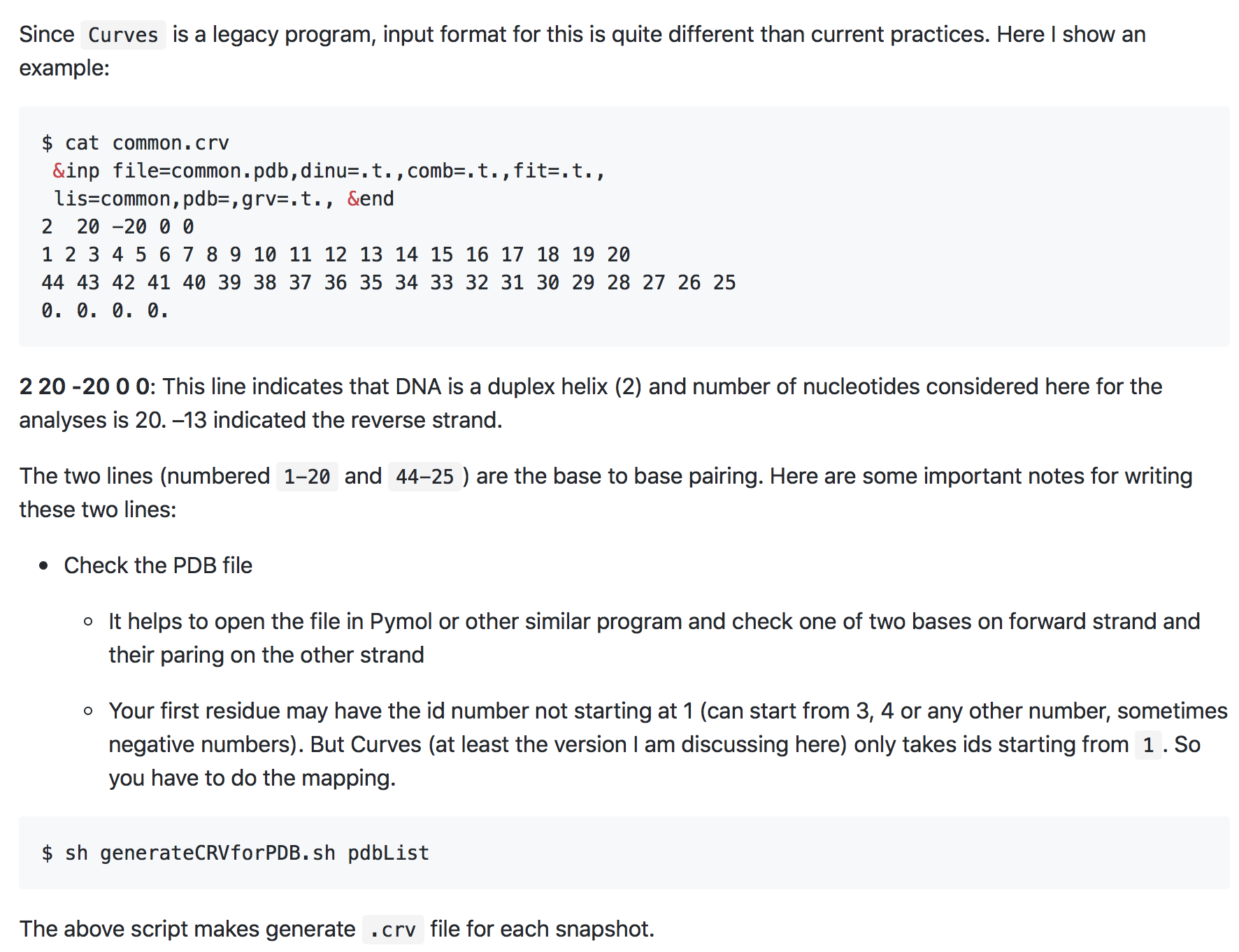
1. Installation of Curves (MacOS or Linux, not for Windows)
   * 1. Copy the file Curves5.3.tar to a directory of your choice in your computer. It does not need to be in the same directory where you process the trajectory.
     2. From a Finder Window, double click on the icon and you will unzip the file, which will produce a folder with all the routines of the program and the executable; the name of the executable is Cur5.
     3. From a Linux or MacOS terminal, give the command “tar –xvf Curves5.3.tar”: this will produce the same effect as in a).
2. Processing the trajectory
   * 1. Go to the folder Trj2shape\_scripts and copy all the .sh and .py scripts to the directory where your MD trajectory is stored (you may not need all the scripts, but some of them are needed indirectly and I am not sure which ones).
     2. Install python on your computer if you don’t have it; I think any python version should work.
     3. The first processing step is to produce many pdb files containing the snapshots of your trajectory; one possibility is to use the gromacs command trjconv; this will generate many pdb files, so it is better if you create a subdirectory of your main directory and copy there the needed .tpr and .xtc files (see image below about this step; the location should be where you have the .tpr and .xtc files and all the scripts in this package; you can do skip 100, so you should get 2000 snapshots for your 200 ns trajectory; # indicates a comment; $ is the Linux prompt).



* + 1. Create a file that contains a list of pdb files on which you wish to run your Curves analysis; see image below; after you give the first command, check the created file pdbList: if it is already in numerical order, then you are set; the second command is to put them in alphabetical order, but the command is OS-specific, you may try the nohup, or you may have to figure out what is the appropriate command.



* + 1. Create a sample input for Curves and generate an input file for each snapshot.

q

* + 1. Run the shape analysis with Curves5.3 for all snapshots: the command is

$ ./curves.sh < pdbList

This command will produce an output with the extension .lis for any snapshot. The script curves.sh contains the command

sed -i back 's:DA: A:;s:DC: C:;s:DG: G:;s:DT: T:g' $pdb  && ~/Documents/structural\_biology\_software/Curves5.3/Cur5 < $crvfile && rm $pdb"back"

Here, you have to specify the path where your Cur5 executable is located, which is where you installed the software.

* + 1. At this point you need to extract relevant information from each .lis file; first, create a file with a list of .lis files

$ find ./ -maxdepth 1 -name "\*.lis" | sort -V | sed 's:.//::g' > lisList

$ awk '{print $1"\t"0"\t"0}' lisList > lisList\_with\_Offset

* + 1. Now you are ready to extract the information

$ python getshape.py -i lisList\_with\_Offset –w finename

You should be able to see four files, filename.{MGW, Roll, HelT, ProT}; you can choose the filename.

Please note that these files include values from found artifacts (see filename.artifact) too. The motivation behind keeping these instances in the shape features files is that you get to see values at each time point. But to get a meaningful average, they should be eliminated.

* + 1. Eliminate artifacts

python fix\_artifact\_issue.py -i filename.MGW -a filename.artifact -o filename.MGW.artifact\_purged

You will obtain a file in which a row corresponds to a snapshot, and for each snapshot you get the values of the minor groove width at each base pair in different columns.

* + 1. Take the average over all the non-artifact snapshots

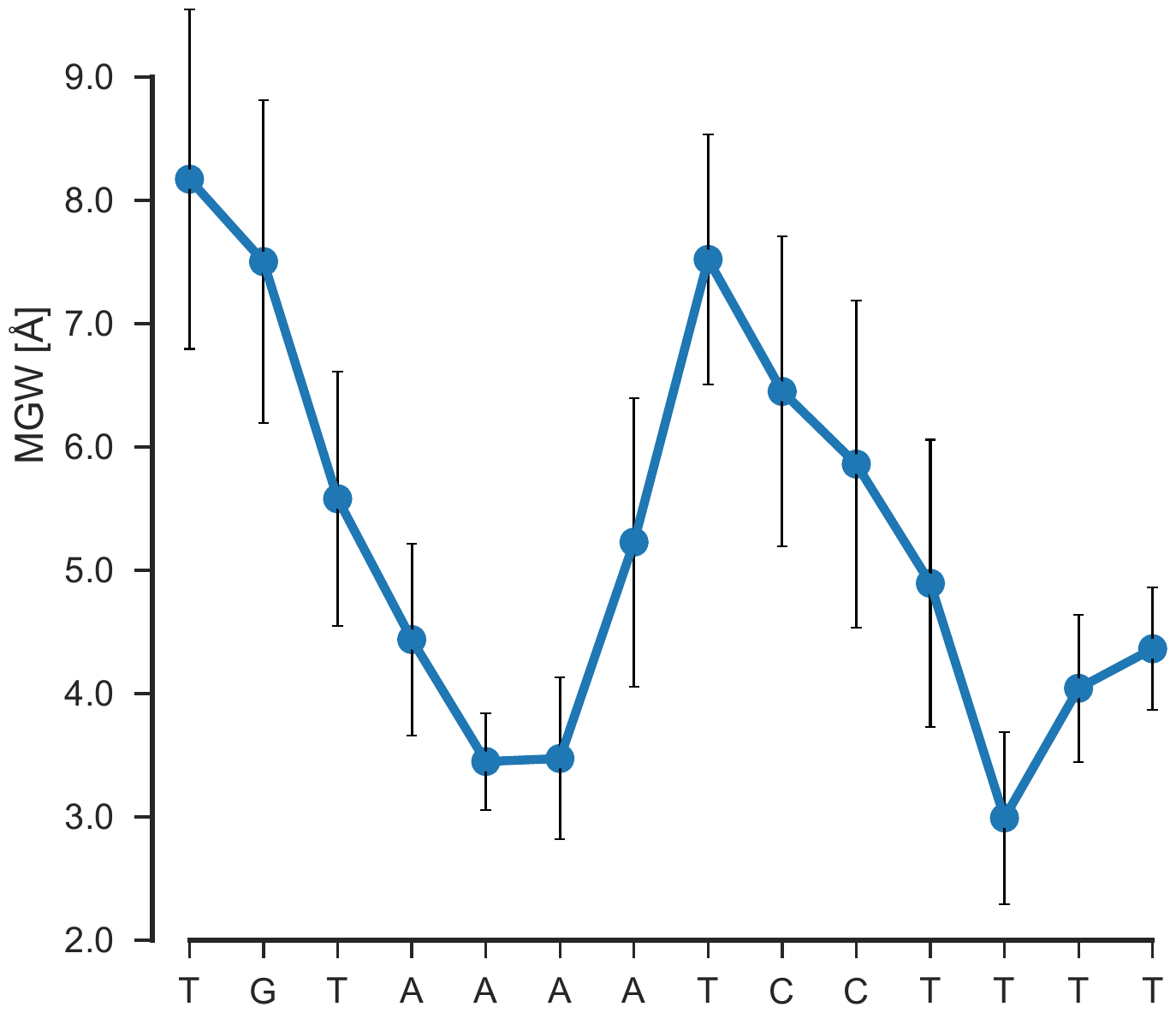
$ python average.py -i filename.MGW.artifact\_purged –o filename.artifact\_purged.avg.MGW

This command should give you a file with 5 columns: column 1 is the base-pair index along your DNA, column 2 is the value of the minor groove width for that bp, column 3 is the standard deviation, column 4 is the number of occurrencies for which you can compute the minor groove width (ideally should be the total number of snapshots, but it is never that, because of the eliminated artifacts), column 5 is the number of missing occurrencies.

At the end you can plot column 2 against column 1 after eliminating base pairs for which you have less than 50% of occurrencies, which normally amounts to 3bp at each end. This will also eliminate the NaN.

The same can be done for the other 3 shape features (Roll, HelT, ProT)

* + 1. Here is a sample plot



* + 1. In the folder DNA\_shape\_parameters\_biblio you find some useful readings about DNA structure and shape parameters