EMBO Course: MD Simulation Trajectory Analysis

Now that we have a trajectory, we can extract some useful information about the system. The first thing we would like to do is to visualize the MD simulation (using a program called VMD). However, before doing this, we need to do some post-processing. In AMBER, all post-processing and analysis is done using the 'ptraj' set of subroutines.

First, we want to remove the translational/rotational motion and calculate the backbone RMSD to the first and average structure. This is done in the following ptraj input files:

ptraj1.inp

```
trajin HPmd.crd
center :1-35
image center familiar
trajout phin.crd
```

ptraj2.inp

```
trajin phin.crd
center :1-35 mass origin
rms first mass out rms_bb_first.dat :1-35@N,C,CA
average phinavg.pdb pdb
trajout phinfit.crd
```

ptraj3.inp

```
trajin phinfit.crd
reference phinavg.pdb.1
rms reference mass out rms_bb_avg.dat :1-35@N,C,CA
trajout phinfitavg.crd
```

Ptraj1.inp allows for the correct visualisation of the MD trajectory under periodic boundary conditions. Ptraj2.inp performs a mass-weighted least-squares fit to the first structure using the N,C' and Cα backbone coordinates for residues 1 to 35, and calculates the average structure. Ptraj3.inp performs a mass-weighted least-squares fit to the average structure. The files *rms_bb_first.dat* and *rms_bb_avg.dat* contain the backbone RMSD in angstrom to the first and average structures respectively. One can plot them in xmgrace.

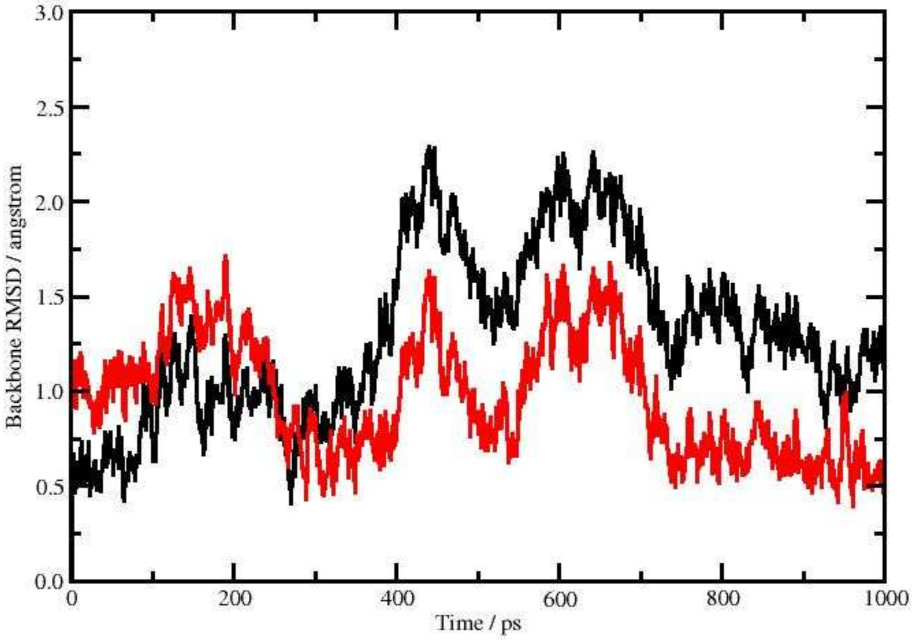
All ptraj commands are run as:

```
ptraj [topology file] [input file]
```

For example:

```
ptraj   HPDRY.parm   ptraj1.inp
```

We can now use VMD to look at the post-processed trajectory.

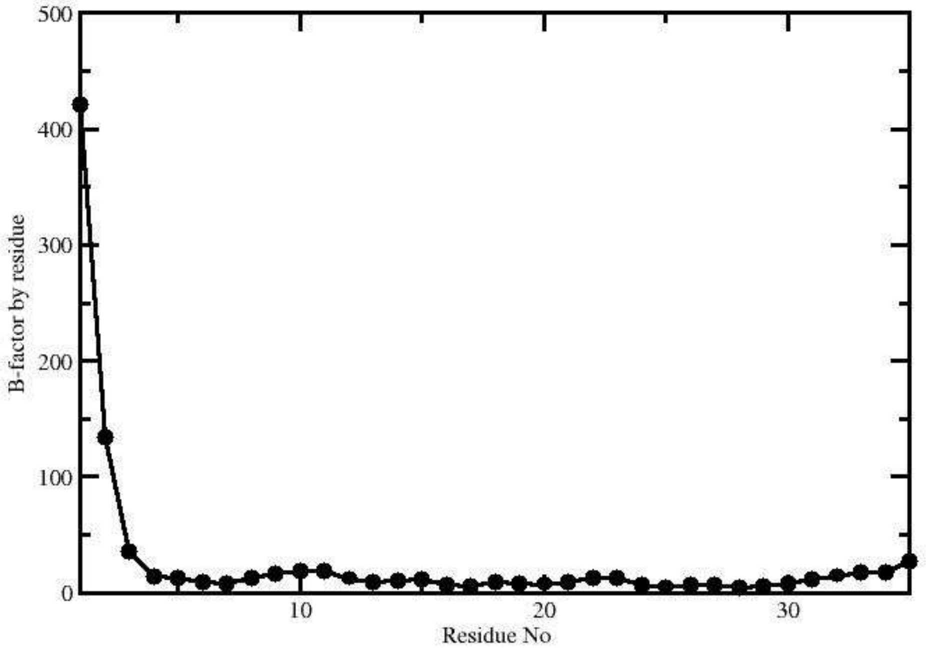


The BB RMSD to the average structure actually gets rather large in this trajectory, increasing to 1.5 angstrom. Proteins typically give BB RMSDs to the average structure between 0.5 and 1 angstrom. In this case, we might have observed a rare event (some large-scale collective conformational change in the protein), or, more likely, some residues in the system, such as those at the termini are very flexible and this causes the BB RMSD value to increase. We could calculate the BB RMSD on a per residue basis, using the trajectory *phinfitavg.crd*, but, it is more appropriate to calculate an experimental observable, such as the B-factor. The B-factor is obtained from X-ray crystallography data and is a measure of the amount of motion for each atom. It can be calculated simply as the atomic fluctuation * (8/3)pi2.**

ptraj4.inp

```
trajin phinfitavg.crd
atomicfluct out backboneBf.dat @C,CA,N byres bfactor
```

Ptraj4 allows us to calculate the B-factor for each residue using the backbone atomic fluctuation.



As can be clearly seen, the B-factor is very large for the first two residues at the N-terminal. This causes the BB RMSD to larger than expected. If we calculated the BB RMSD without the first two residues, the value would be much smaller.

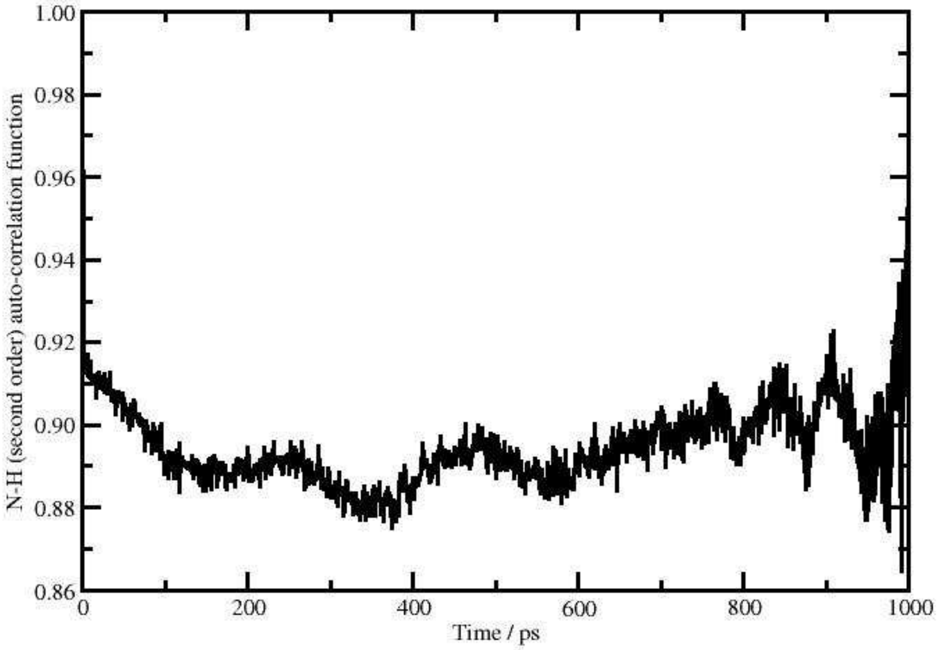
An alternative experimental approach to quantify the amount of dynamics present in bio-molecules is NMR-based spin relaxation: Inducing magnetisation (a large magnetic dipole) on a nucleus in the presence of an external magnetic field, brings the system away from equilibrium. Over time, the system returns to equilibrium (the magnetisation on the nucleus is dissipated through the molecule), a process called *'relaxation'*. The relaxation is mediated by magnetic dipole-dipole couplings (interactions) between spin active nuclei described by a uni-axial second rank interaction tensor, with the principal axis pointing directly along the inter-nuclear vector. The relaxation rate is determined by the dynamics of the system. The most common spin relaxation mechanism that is measured experimentally is N-H dipole-dipole auto-relaxation. The following ptraj script calculates the re-orientational correlation function for the N-H dipole-dipole interaction tensor. The auto-correlation function represents how the system (the second rank tensor) looses memory of its state over time.

ptraj5

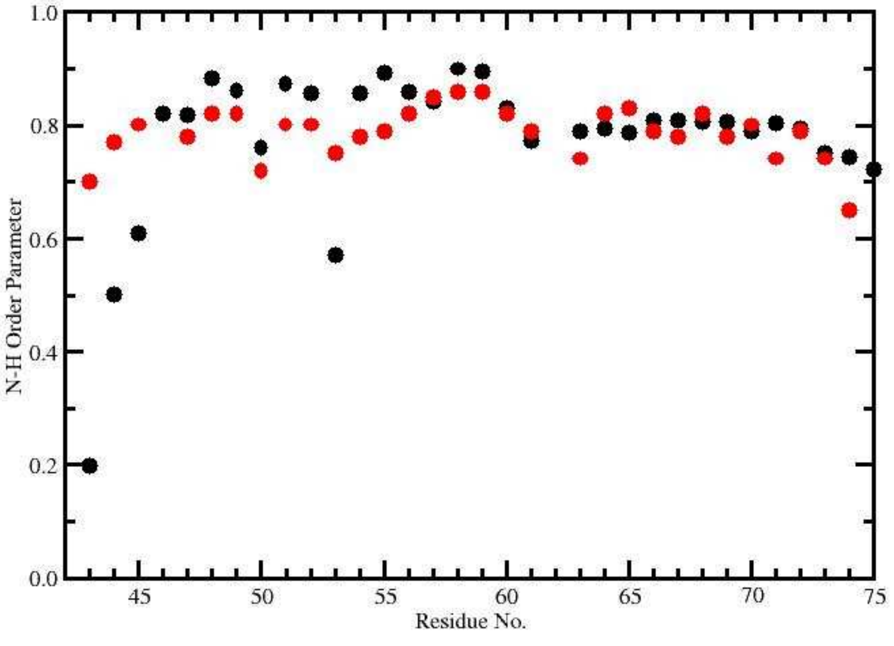
```
trajin phinfitavg.crd
vector v2  @22  corr @23  order 2
vector v3  @33  corr @34  order 2
vector v4  @45  corr @46  order 2
vector v5  @60  corr @61  order 2
vector v6  @72  corr @73  order 2
vector v7  @92  corr @93  order 2
vector v8  @114 corr @115 order 2
vector v9  @124 corr @125 order 2
vector v10 @140 corr @141 order 2
vector v11 @160 corr @161 order 2
vector v12 @167 corr @168 order 2
vector v13 @184 corr @185 order 2
vector v14 @198 corr @199 order 2
```

```
vector v15 @222 corr @223 order 2
vector v16 @233 corr @234 order 2
vector v17 @243 corr @244 order 2
vector v18 @263 corr @264 order 2
vector v19 @273 corr @274 order 2
vector v20 @287 corr @288 order 2
vector v22 @320 corr @321 order 2
vector v23 @339 corr @340 order 2
vector v24 @363 corr @364 order 2
vector v25 @385 corr @386 order 2
vector v26 @402 corr @403 order 2
vector v27 @419 corr @420 order 2
vector v28 @433 corr @434 order 2
vector v29 @452 corr @453 order 2
vector v30 @474 corr @475 order 2
vector v31 @496 corr @497 order 2
vector v32 @511 corr @512 order 2
vector v33 @533 corr @534 order 2
vector v34 @540 corr @541 order 2
vector v35 @559 corr @560 order 2
analyze timecorr vec1 v2  vec2 v2 tstep 1.0 tcorr 1000.0 out corr42.out
analyze timecorr vec1 v3  vec2 v3 tstep 1.0 tcorr 1000.0 out corr43.out
analyze timecorr vec1 v4  vec2 v4 tstep 1.0 tcorr 1000.0 out corr44.out
analyze timecorr vec1 v5  vec2 v5 tstep 1.0 tcorr 1000.0 out corr45.out
analyze timecorr vec1 v6  vec2 v6 tstep 1.0 tcorr 1000.0 out corr46.out
analyze timecorr vec1 v7  vec2 v7 tstep 1.0 tcorr 1000.0 out corr47.out
analyze timecorr vec1 v8  vec2 v8 tstep 1.0 tcorr 1000.0 out corr48.out
analyze timecorr vec1 v9  vec2 v9 tstep 1.0 tcorr 1000.0 out corr49.out
analyze timecorr vec1 v10 vec2 v10 tstep 1.0 tcorr 1000.0 out corr50.out
analyze timecorr vec1 v11 vec2 v11 tstep 1.0 tcorr 1000.0 out corr51.out
analyze timecorr vec1 v12 vec2 v12 tstep 1.0 tcorr 1000.0 out corr52.out
analyze timecorr vec1 v13 vec2 v13 tstep 1.0 tcorr 1000.0 out corr53.out
analyze timecorr vec1 v14 vec2 v14 tstep 1.0 tcorr 1000.0 out corr54.out
analyze timecorr vec1 v15 vec2 v15 tstep 1.0 tcorr 1000.0 out corr55.out
analyze timecorr vec1 v16 vec2 v16 tstep 1.0 tcorr 1000.0 out corr56.out
analyze timecorr vec1 v17 vec2 v17 tstep 1.0 tcorr 1000.0 out corr57.out
analyze timecorr vec1 v18 vec2 v18 tstep 1.0 tcorr 1000.0 out corr58.out
analyze timecorr vec1 v19 vec2 v19 tstep 1.0 tcorr 1000.0 out corr59.out
analyze timecorr vec1 v20 vec2 v20 tstep 1.0 tcorr 1000.0 out corr60.out
analyze timecorr vec1 v22 vec2 v22 tstep 1.0 tcorr 1000.0 out corr62.out
analyze timecorr vec1 v23 vec2 v23 tstep 1.0 tcorr 1000.0 out corr63.out
analyze timecorr vec1 v24 vec2 v24 tstep 1.0 tcorr 1000.0 out corr64.out
analyze timecorr vec1 v25 vec2 v25 tstep 1.0 tcorr 1000.0 out corr65.out
analyze timecorr vec1 v26 vec2 v26 tstep 1.0 tcorr 1000.0 out corr66.out
analyze timecorr vec1 v27 vec2 v27 tstep 1.0 tcorr 1000.0 out corr67.out
analyze timecorr vec1 v28 vec2 v28 tstep 1.0 tcorr 1000.0 out corr68.out
analyze timecorr vec1 v29 vec2 v29 tstep 1.0 tcorr 1000.0 out corr69.out
analyze timecorr vec1 v30 vec2 v30 tstep 1.0 tcorr 1000.0 out corr70.out
analyze timecorr vec1 v31 vec2 v31 tstep 1.0 tcorr 1000.0 out corr71.out
analyze timecorr vec1 v32 vec2 v32 tstep 1.0 tcorr 1000.0 out corr72.out
analyze timecorr vec1 v33 vec2 v33 tstep 1.0 tcorr 1000.0 out corr73.out
analyze timecorr vec1 v34 vec2 v34 tstep 1.0 tcorr 1000.0 out corr74.out
analyze timecorr vec1 v35 vec2 v35 tstep 1.0 tcorr 1000.0 out corr75.out
```

For each backbone N-H moiety we now have an auto-correlation function describing the internal motion of the N-H bond vector. As an example, for residue 54(THR), the re-orientational (second order) auto-correlation function looks like this:



At time t = 0, the auto-correlation function is 1, and the system has an exact 'memory' of its state. In the first 2 pico-seconds, the correlation function decays rapidly to about 0.915, representative of the fast backbone dihedral angle fluctuations of the residue. Over the next several hundred pico-seconds, the correlation function decays to a plateau at about 0.885. The value at this plateau is called the order parameter (S^2) and provides a quantitative measure of the amount of internal re-orientational motion. The larger the order parameter, the more rigid the system is. Order parameters can be extracted from NMR spin relaxation experiments. A simple way of calculating the order parameter is to perform a summation over direct cosines. A FORTRAN code is available to do this. The order parameters can be compared to available experimental data.

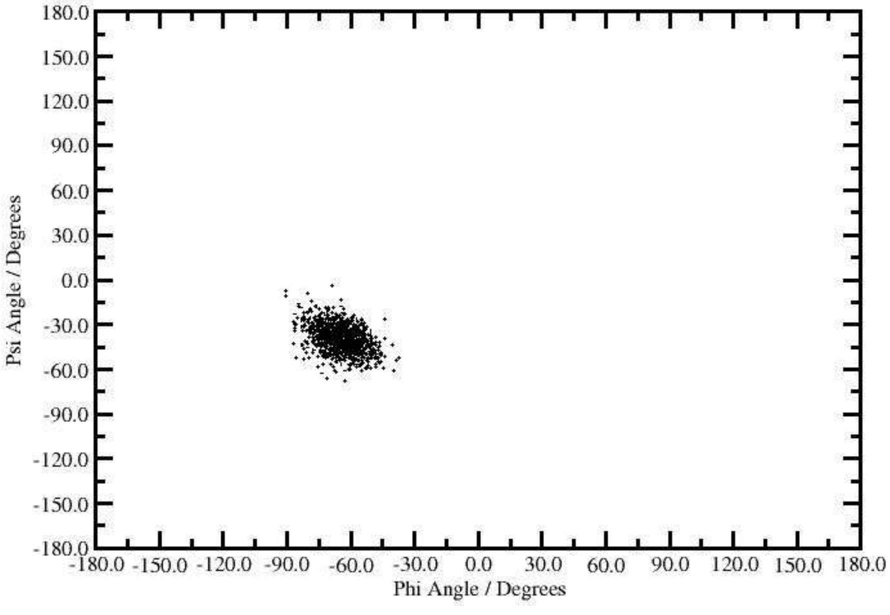


The agreement between experiment (red circles) and theory (black circles) is rather good, particularly considering that this is a single 1 ns trajectory. The N-terminal residues show far too much motion. This is due to the fact that we chopped off the N-terminal part of the molecule and only performed a simulation on the headpiece. The only other outsider is residue 53(MET). Notice that the order parameters here are lower than the plateau values in the auto-correlation function. This is because the correlation functions only treat re-orientational motion, however, the relaxation is also mediated by the vibrational motion of the N-H bond. The vibrational component can be introduced using a scaling factor of 0.89. Averaging the results over multiple trajectories (using different random seed generators) will lead to further improvement.

Finally, we can also look at the backbone dihedral angles (ϕ,φ) for each residue, and generate a Ramachandran plot. The ptraj input file to calculate the phi/psi angles for residue 30 looks like this:

ptraj6.inp

```
trajin phinfitavg.crd
dihedral 30psi :30@N :30@CA :30@C :31@N out psi30.dat
dihedral 30phi :29@C :30@N  :30@CA :30@C out phi30.dat
```



The villan headpiece is comprised of three α-helices. The Ramachandran plot confirms that the backbone geometry of this residue is indeed that of an α-helix.

In principal, there is no end to the amount of analysis one can perform as all the structural/dynamic information is available. This short set of examples represents just a few possibilities.