p16_trajectory_analysis

Tutorial 2

Comparative analysis of structural changes of the P16 protein (CDKN2A gene) wild type (WT) with mutants:

1. Analysis using basic MD metrics — RMSD, Rg, RMSF :

After the initial relaxing 200 ns production run, we evaluate the protein trajectory using the basic MD analysis metrics:

1.1 RMSD (оценка стабильности структуры во времени)

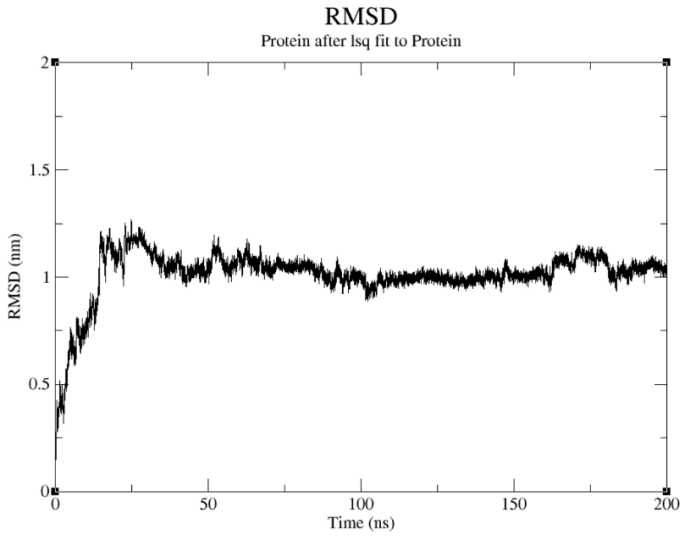
```
gmx rms -s md_200ns.tpr -f md_200ns_fit.xtc -o rmsd.xvg -tu ns
```

Groups:
Fit group → `1(Protein)`
RMSD group → `1(Protein)`

-s md_200ns.tpr — *input tpr file (reference structure + parameters)*
-f md_200ns_fit.xtc — *trajectory after centering and PBC correction*
-o rmsd.xvg — *output file with RMSD over time*
-tu ns — *output time in nanoseconds (ns) instead of ps*

Output file:
rmsd.xvg — *RMSD(t) plot over the entire 200 ns trajectory.*

Plot the .xvg graphs using any convenient tool (xmgrace, Python, Excel, etc.):



Output the statistics:
`gmx analyze -f rmsd.xvg`

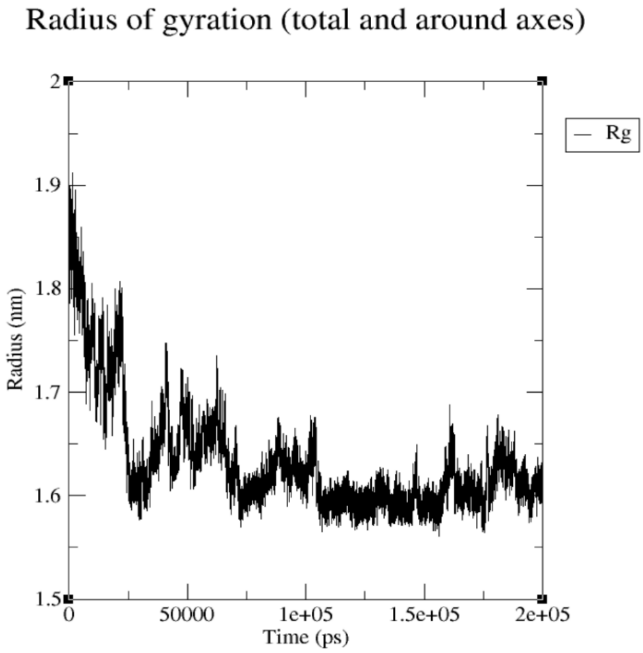
```
                                    std. dev.    relative deviation of
                      standard      ---------   cumulants from those of
set      average      deviation     sqrt(n-1)   a Gaussian distribition
                                                   cum. 3   cum. 4
SS1   1.012291e+00   1.202844e-01   3.803728e-04     -1.948    4.442
```

1.2 Radius of Gyration (Rg)

```
gmx gyrate -s md_200ns.tpr -f md_200ns_fit.xtc -o gyrate.xvg
```

Group: `Protein`

Output file:
gyrate.xvg — *change in protein compactness (Rg) over time.*



Statistics:
`gmx analyze -f gyrate.xvg`

set	average	standard deviation	std. dev. / sqrt(n-1)	relative deviation of cumulants from those of a Gaussian distribition: cum. 3	cum. 4
SS1	1.633173e+00	5.430302e-02	1.717212e-04	1.202	1.243
SS2	1.322463e+00	1.322524e-01	4.182189e-04	-0.220	-0.272
SS3	1.318644e+00	1.201321e-01	3.798912e-04	-0.246	-0.068
SS4	1.344767e+00	1.111679e-01	3.515439e-04	-0.142	0.008

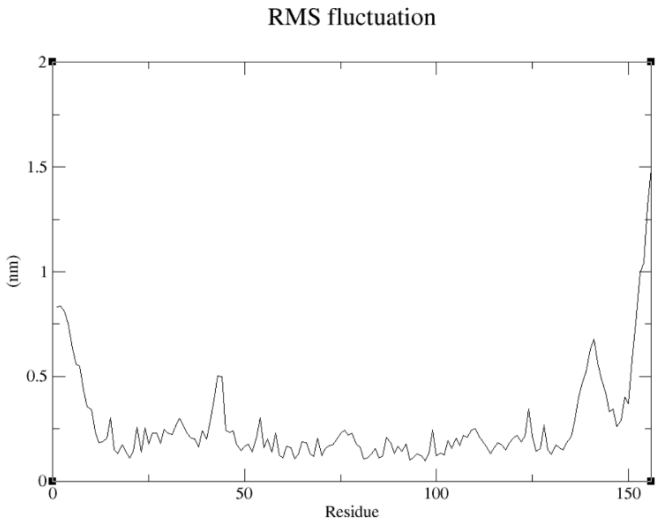
**SS1 ≡ total radius of gyration (Rg_total), SS2, SS3, SS4 — components of Rg along the axes (X, Y, Z).*

1.3 RMSF (per-residue fluctuations)

```
gmx rmsf -s md_200ns.tpr -f md_200ns_fit.xtc -o rmsf.xvg -res
```

Select the group: `Protein`

Output file:
rmsf.xvg — *RMSF for each protein residue (156 aa).*



Statistics:
`gmx analyze -f rmsf_WT.xvg`

set	average	standard deviation	std. dev. / sqrt(n-1)	relative deviation of cumulants from those of a Gaussian distribition cum. 3	cum. 4
SS1	2.722282e-01	2.189509e-01	1.758656e-02	1.783	3.173

Comparative analysis

As the wild-type (WT) template, we use the last frame (PDB) of the initial 200 ns production molecular dynamics run, taken from the centered trajectory(.xtc).

The resulting structure is cleaned of water molecules and ions, after which it is used as the starting structure for performing in silico mutagenesis (using third-party software, e.g., ChimeraX, Maestro Schrödinger).
*(Files: **p16_p48l_h83y.pdb**, **p16_r131p_g55c.pdb**)*

The mutations selected for comparison with WT correspond to real clinical cases and were chosen based on patient data available on cBioPortal — an international platform for the analysis of cancer genomics data.

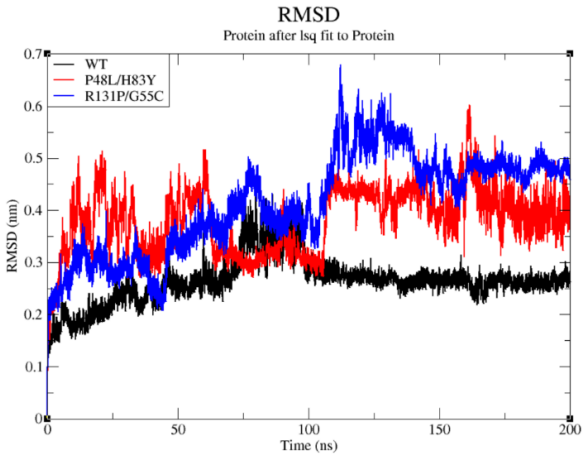
Mutations:
P48L + H83Y: *https://www.cbioportal.org/patient?sampleId=P-0006661-T01-IM5&studyId=pdac_msk_2024*
R131P + G55C: *https://www.cbioportal.org/patient?sampleId=P-0003222-T01-IM5&studyId=lung_msk_2017*

Both the mutants and the WT undergo the standard pipeline followed by a 200 ns production run (as described in Tutorial 1).

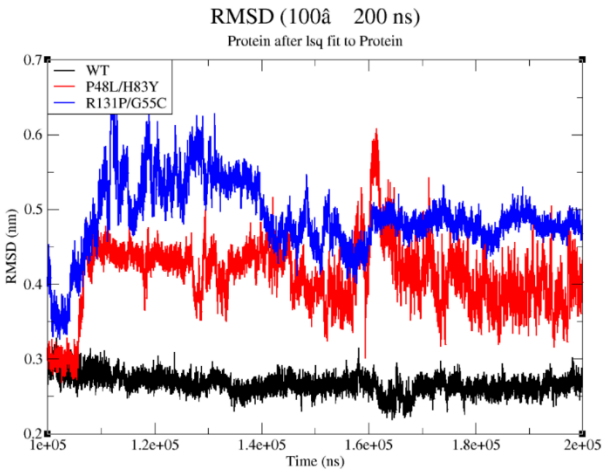
2.1 Comparison using basic metrics (RMSD, Rg, RMSF):

Use the commands described in the previous section (`gmx rms, gmx gyrate, gmx rmsf`).
Combine the plots (the .xvg files) using any convenient method (xmgrace, Python, Excel, etc.).

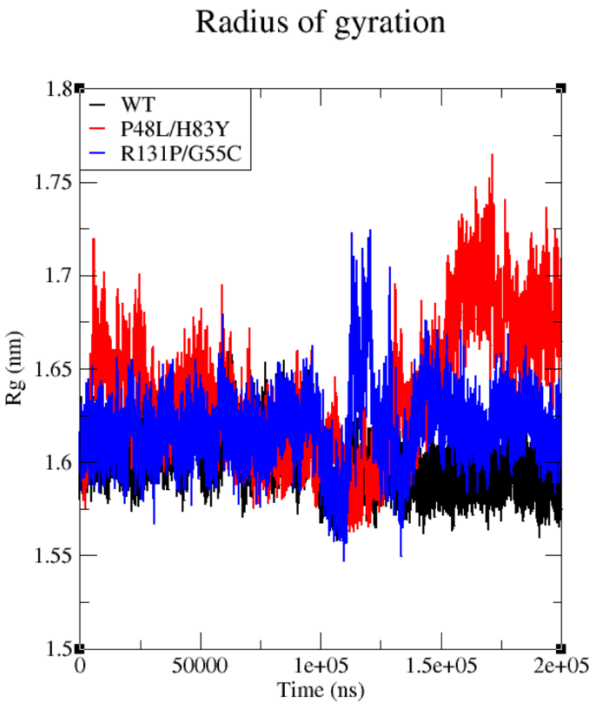
RMSD — the 100–200 ns segment was calculated separately (second plot)

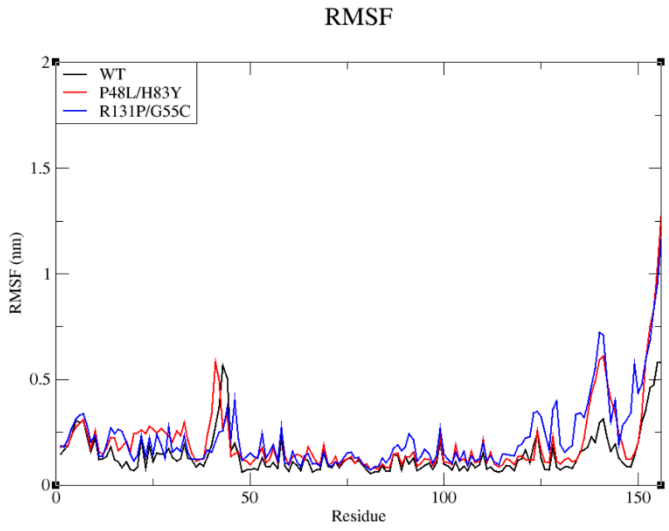


0-200ns



Time interval 100–200 ns





Using `gmx analyze` *collect the data into a similar table for further comparison:*

Белок	RMSD Avg	RMSD Std	RMSF Avg	RMSF Std	Rg Avg	Rg Std
WT						
P48L/H83Y						
R131P/G55C						

Result:

Белок	RMSD Avg	RMSD Std	RMSD Avg (100–200нс)	RMSD Std (100–200нс)	RMSF Avg	RMSF Std	Rg Avg	Rg Std
WT	0.263	0.045	**0.266**	**0.012**	0.145	0.100	1.600	0.014
P48L/H83Y	0.380	0.062	**0.411**	**0.047**	0.204	0.167	1.637	0.034
R131P/G55C	0.412	0.096	**0.490**	**0.048**	0.225	0.164	1.618	0.019

Both mutations (P48L/H83Y and R131P/G55C) lead to increased dynamic mobility of the protein compared to WT, which is reflected in the rise of the average RMSD from ~0.26 nm (WT) to ~0.38 nm (P48L/H83Y) and ~0.41 nm (R131P/G55C) over the entire simulation interval (0–200 ns). The mutants also exhibit a slightly larger radius of gyration (Rg ≈ 1.64 nm) compared to WT (~1.60 nm), indicating a more "loosened" conformation.

In the 100–200 ns interval, the differences become more pronounced: RMSD remains low for WT (~0.27 nm), whereas it increases to ~0.41 nm for P48L/H83Y and reaches ~0.49 nm for R131P/G55C. This confirms greater structural deviation and increased flexibility in the mutants, especially in the case of R131P/G55C.

2.2 Protein Energy:

```
gmx energy -f WT.edr -o WT_energy.xvg
```

```
Energy                    Average   Err.Est.       RMSD  Tot-Drift
-------------------------------------------------------------------------------
Potential                -532540         14    900.967   -85.2964  (kJ/mol)
Temperature              300.006      0.002    1.62805 -0.00654826  (K)
Pressure                 1.08504       0.11    149.284  -0.249805  (bar)
```

```
gmx energy -f P48L_H83Y.edr -o P48L_H83Y_energy.xvg
```

```
Energy                    Average   Err.Est.       RMSD  Tot-Drift
-------------------------------------------------------------------------------
Potential                -532691         15    901.609   -76.2976  (kJ/mol)
Temperature              300.007     0.0035    1.62961 -0.0097168  (K)
Pressure                 1.07503      0.055    149.744  -0.193645  (bar)
```

```
gmx energy -f R131P_G55C.edr -o R131P_G55C_energy.xvg
```

```
Energy                    Average   Err.Est.       RMSD  Tot-Drift
-------------------------------------------------------------------------------
Potential                -531972          8.1    902.473    19.4028  (kJ/mol)
Temperature              300.005     0.0023    1.62851 0.00910602  (K)
Pressure                 1.03275       0.07    150.139   0.381466  (bar)
```

- ***Temperature*** *should fluctuate around the target value (e.g., 300 K) with RMSD ~1–5 K and without noticeable drift.*
- ***Pressure in the NPT*** *ensemble exhibits large fluctuations (RMSD ~50–500 bar); evaluation is based on the average value, which should be close to 1 bar.*
- ***Potential energy*** *should be stable, fluctuating around a mean value with minimal drift relative to |E|.*
- ***Tot-Drift*** *should be small; significant values indicate numerical instability of the simulation.*

2.3 Intra-protein H-bonds:

Look at the average number of hydrogen bonds over the entire trajectory

Example command (WT):

```
gmx hbond -s WT.tpr -f md_200ns_fit.xtc -num WT_hbnum_PP.xvg -WT_hbonds_PP.ndx
```

-s WT.tpr — *input tpr file with system topology and parameters*
-f md_200ns_fit.xtc — *trajectory*
-num WT_hbnum_PP.xvg — *output number of hydrogen bonds over time (Protein–Protein)*
-o WT_hbonds_PP.ndx — *save the list of all donor–acceptor pairs (index file)*

Select: `group "Protein"`

Output files:
WT_hbnum_PP.xvg — *time evolution of the number of Protein–Protein hydrogen bonds (time → H-bond count)*
WT_hbonds_PP.ndx — *complete list of donor–acceptor pairs within the protein*

Result check:

```
gmx analyze -f WT_hbnum_PP.xvg
```

```
Command line:
  gmx analyze -f WT_hbnum_PP.xvg

Read 1 sets of 100001 points, dt = 2

                                    std. dev.    relative deviation of
                      standard      ---------   cumulants from those of
set      average      deviation     sqrt(n-1)   a Gaussian distribition
                                                   cum. 3   cum. 4
SS1   1.093712e+02   5.513737e+00   1.743597e-02     -0.066    0.020
```

Average number of intra-protein hydrogen bonds:

Protein system	N(H-bond)
WT	**109.37**
P48L_H83Y	**109.59**
R131P_G55C	**107.03**

In terms of the number of intramolecular hydrogen bonds, the P48L_H83Y mutant is practically indistinguishable from WT, whereas the R131P_G55C mutant shows a slight decrease (~2%), which may reflect weakening of the protein's internal packing.

2.4 DSSP analysis (secondary structure):

** **mkdssp** must be pre-installed*

Example command (WT):

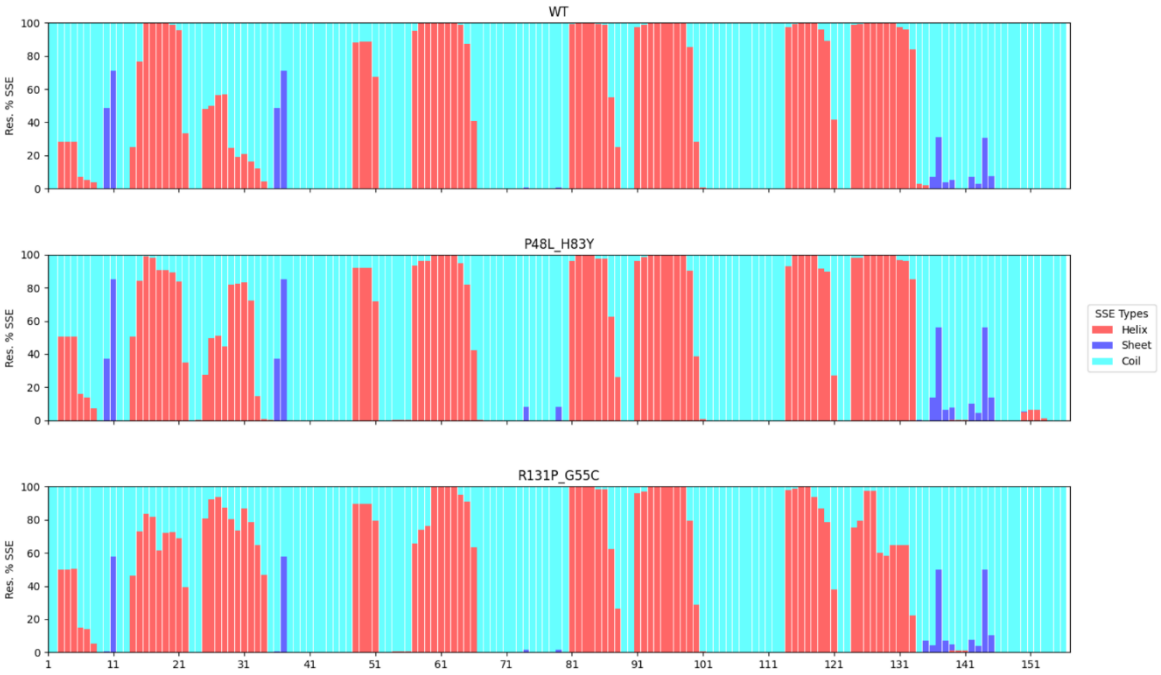
```
gmx dssp -s WT.tpr -f md_200ns_fit.xtc \
  -o WT_dssp.dat -num WT_dssp_num.xvg \
  -sel 'group "Protein"'
```

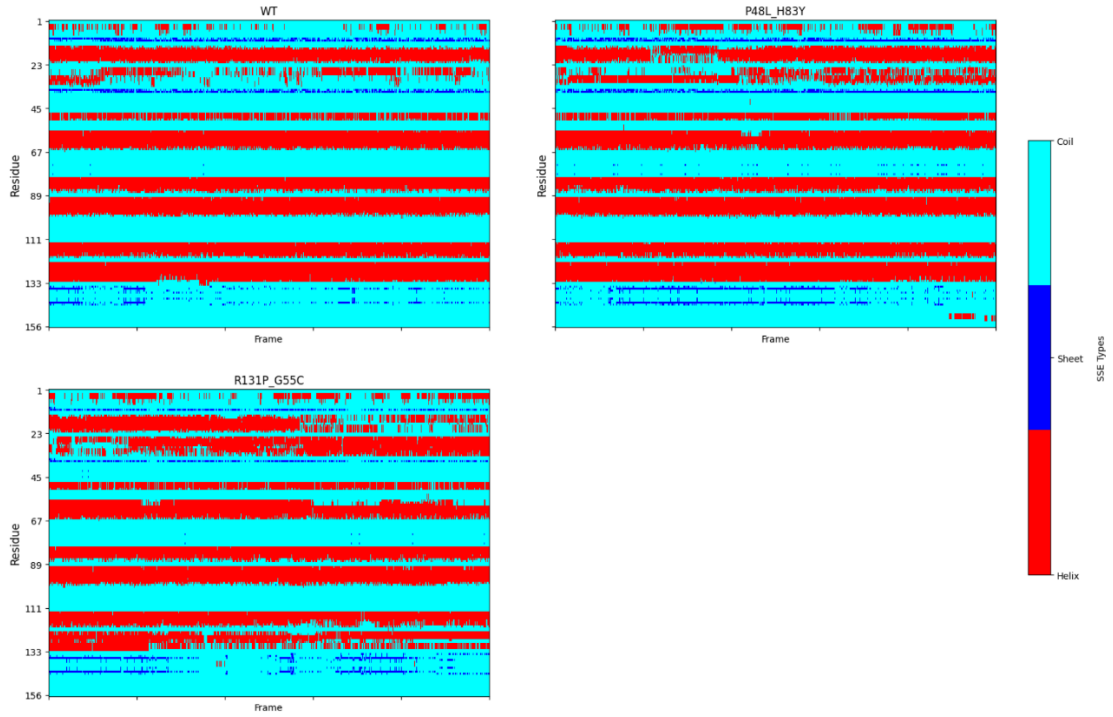
gmx dssp — *calculation of protein secondary structure (DSSP)*
-s WT.tpr — *system topology*
-f md_200ns_fit.xtc — *centered trajectory without PBC artifacts*
-o WT_dssp.dat — *DSSP annotation by frames and residues*
-num WT_dssp_num.xvg — *numerical statistics of H/E/C over time*
-sel 'group "Protein"' — *analyze protein only*

Output:
WT_dssp.dat — *secondary structure assignment for analysis and plotting*
WT_dssp_num.xvg — *control statistics*

For analysis and visualization of the protein secondary structure over time, you can use the script* **sse_P16.ipynb

WT - Helix: 36.41%, Sheet: 2.15%, Coil: 61.43%
P48L_H83Y - Helix: 38.26%, Sheet: 2.75%, Coil: 58.98%
R131P_G55C - Helix: 36.84%, Sheet: 1.69%, Coil: 61.47%





2.5 PCA (Principal Component Analysis) = analysis of the principal components of motion.

Compute the covariance matrix of atomic positional fluctuations → calculate its eigenvectors (PC1, PC2, PC3 …).

Example command (WT):

```
gmx covar \
  -s WT.tpr \
  -f md_200ns_fit.xtc \
  -o WT_eigenval.xvg \
  -v WT_eigenvec.trr \
  -av WT_average.pdb \
  -b 100000
```

Select a group: 4
Selected 4: 'Backbone'

-s WT.tpr — *system topology*
-f md_200ns_fit.xtc — *centered trajectory without PBC artifacts*
-o WT_eigenval.xvg — *eigenvalues (contribution of each mode)*
-v WT_eigenvec.trr — *eigenvectors (principal motions)*
-av WT_average.pdb — *average protein structure*
-b 100000 — *start analysis from 100 ns (initial relaxation is discarded)*

Output:
WT_eigenval.xvg — *mode spectrum (PCA)*
WT_eigenvec.trr — *trajectory of principal components*
WT_average.pdb — *average structure*

PCA 1 vs 2 :

This is the projection of the trajectory onto the first and second principal components (PC1 and PC2). It shows the main conformational landscape.

Example command (WT):

```
gmx anaeig \
  -v WT_eigenvec.trr \
  -s WT.tpr \
  -f md_200ns_fit.xtc \
  -2d WT_2D_PC1_PC2.xvg \
  -b 100000 \
  -first 1 \
  -last 2
```

```
Select a group: 4/4
Selected 4: 'Backbone'
```

-v WT_eigenvec.trr — *eigenvectors from gmx covar*
-s WT.tpr — *system topology*
-f md_200ns_fit.xtc — *centered trajectory*
-2d WT_2D_PC1_PC2.xvg — *projection of the trajectory onto PC1–PC2*
-b 100000 — *start analysis from 100 ns*
-first 1 — *first principal component (PC1)*
-last 2 — *second principal component (PC2)*

Output:
WT_2D_PC1_PC2.xvg — *2D conformational landscape (PC1 vs PC2)*

PCA 1 vs 2 :

This is the projection of the trajectory onto the first and second principal components (PC1 and PC2). It shows the main conformational landscape.

Example command (WT):

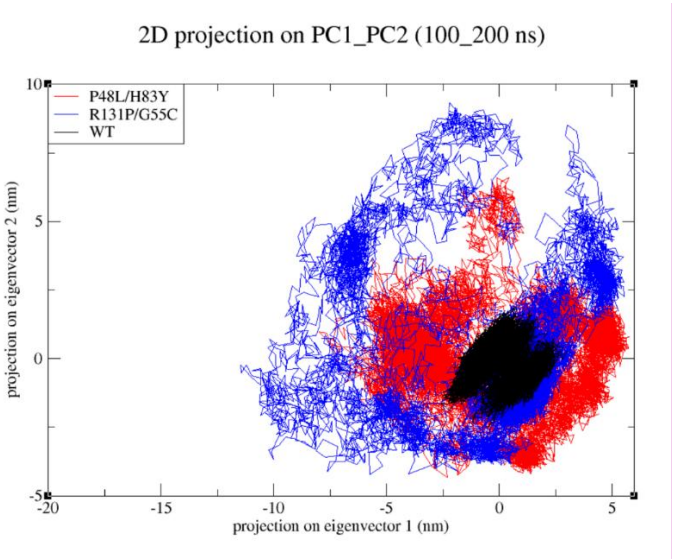
```
gmx anaeig \
  -v WT_eigenvec.trr \
  -s WT.tpr \
  -f md_200ns_fit.xtc \
  -2d WT_2D_PC1_PC2.xvg \
  -b 100000 \
  -first 1 \
  -last 2
```

```
Select a group: 4/4
Selected 4: 'Backbone'
```

-v WT_eigenvec.trr — *eigenvectors from gmx covar*
-s WT.tpr — *system topology*
-f md_200ns_fit.xtc — *centered trajectory*
-2d WT_2D_PC1_PC2.xvg — *projection of the trajectory onto PC1–PC2*
-b 100000 — *start analysis from 100 ns*
-first 1 — *first principal component (PC1)*
-last 2 — *second principal component (PC2)*

Output:
WT_2D_PC1_PC2.xvg — *2D conformational landscape (PC1 vs PC2)*

Plot the graphs. They can be overlaid on top of each other



PCA 1 vs 3 :

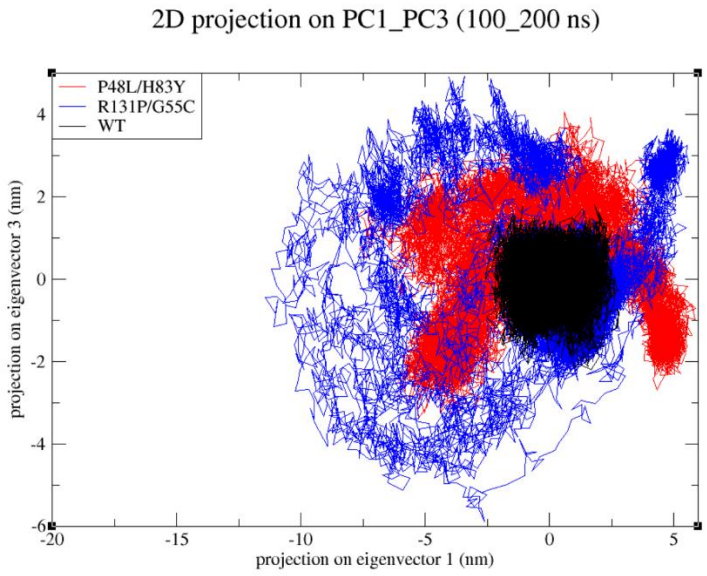
Reveals additional / secondary motions that may not be visible in PC1–PC2 (alternative PCA projection).

Example command (WT):

```
gmx anaeig \
  -v WT_eigenvec.trr \
  -s WT.tpr \
  -f md_200ns_fit.xtc \
  -2d WT_2D_PC1_PC3.xvg \
  -b 100000 \
  -first 1 \
  -last 3
```

```
Select a group: 4/4
Selected 4: 'Backbone'
```

Plot the graphs. They can be overlaid on top of each other



2.6 ProLif is a library for analyzing interactions in MD trajectories.

Analysis of MD trajectories that shows which amino acids interact with each other, what types of contacts they form, and how often over time; it is needed to quantitatively compare interactions in WT and mutants and understand which structural contacts are preserved, lost, or newly formed.

Detailed tutorial here: https://prolif.readthedocs.io/en/stable/source/tutorials.html
Example calculation for our trajectories - **P16_prolif.ipynb**

[16]:

	total_unique_interactions	unique_interactions_per_frame
WT	2948416.0	736.919770
P48L_H83Y	2931819.0	732.771557
R131P_G55C	2870260.0	717.385654

17]:

interaction	WT	P48L_H83Y	R131P_G55C
Anionic	57151	57107	52689
Cationic	57151	57107	52689
HBAcceptor	397762	403444	392518
HBDonor	389398	390589	374899
Hydrophobic	569589	551157	529601
PiCation	4367	4146	3236
VdWContact	1472998	1468269	1464628

P48L_H83Y *— minimal differences compared to WT: a slight decrease in hydrophobic contacts is compensated by preservation or a slight increase in hydrogen bonding, indicating local rearrangement without global destabilization.*

R131P_G55C *— shows a noticeable reduction in hydrophobic, ionic, and π–cation contacts, suggesting disruption of tight packing and weakening of stabilizing interactions.*
Both mutants demonstrate a reduction in the number of intra-protein interactions compared to WT, with the effect being more pronounced for R131P_G55C, indicating a more substantial weakening of structural integrity.

2.7 Visual analysis of trajectories

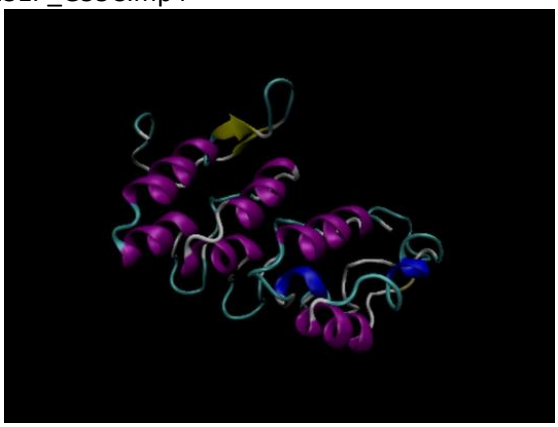
For visualization, we will use ***VMD.***

After completing the MD simulation, first load the structure, then overlay the trajectory. To do this, launch VMD, open ***File → New Molecule****, and load the structure file (.gro). Then, in the same window, click* ***Load Files*** *and add the trajectory file (.xtc). The trajectory will automatically play as an animation over the original structure.*

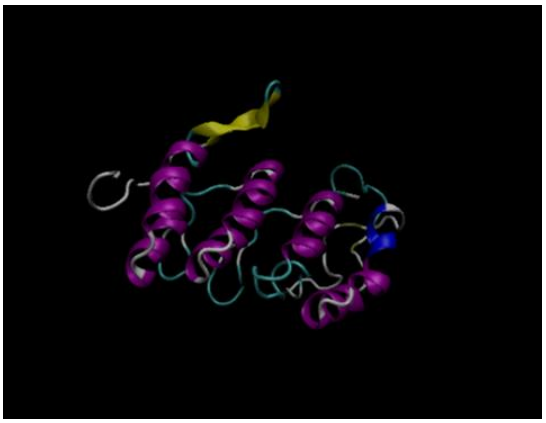
In VMD, the display style is configured via ***Graphics → Representations****: typically, a representation is created with the selection* `protein` *and the* ***NewCartoon*** *drawing method enabled, colored by secondary structure. Water can either be completely hidden by setting the selection to* `not water`*, or displayed separately in a simplified form (e.g.,* ***Lines*** *or* ***Points****) to avoid cluttering the visualization.*

The visualization can be recorded using standard screen recording tools.

P16_WT.mp4
P16_R131P_G55C.mp4



WT



R131P + G55C

Official VMD tutorials:
https://www.ks.uiuc.edu/Training/Tutorials/vmd/tutorial-html/
https://www.ks.uiuc.edu/Training/Tutorials/vmd-imgmv/imgmv/imgmv-tutorial.pdf