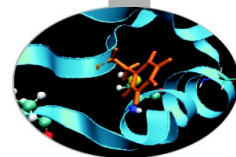


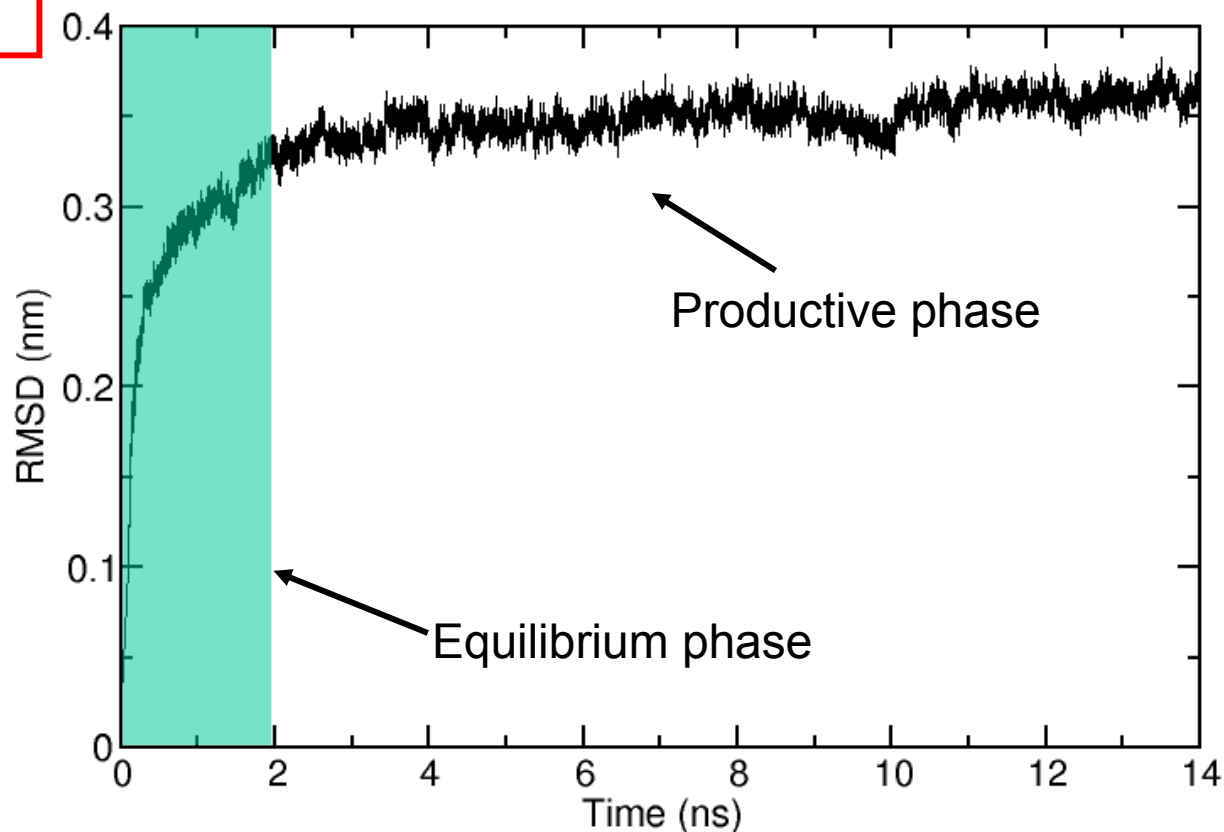
Analysis of MD trajectories (Essential Dynamics of Proteins)

Did we reach equilibrium...?

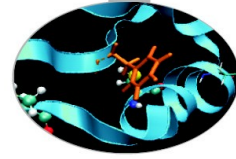


$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - r_0)^2}$$

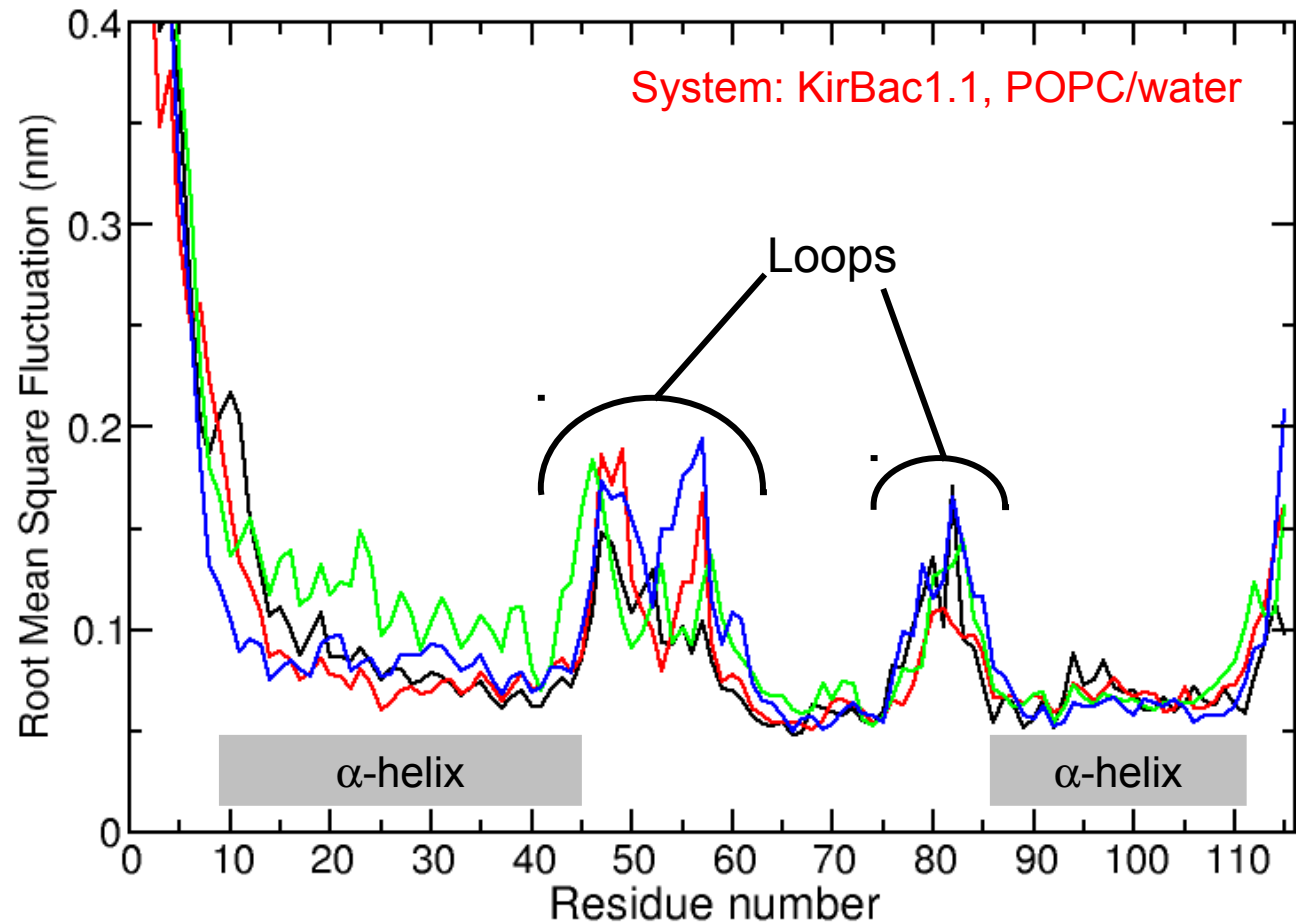
We need to make sure that all the chemical and physical properties of the system have reached an equilibrium, where their averages do not longer change as a function of time. A simple way to test this is by measuring the RMSD (root mean square deviation) of C α carbon atoms position with respect to start.



Measuring chain flexibility

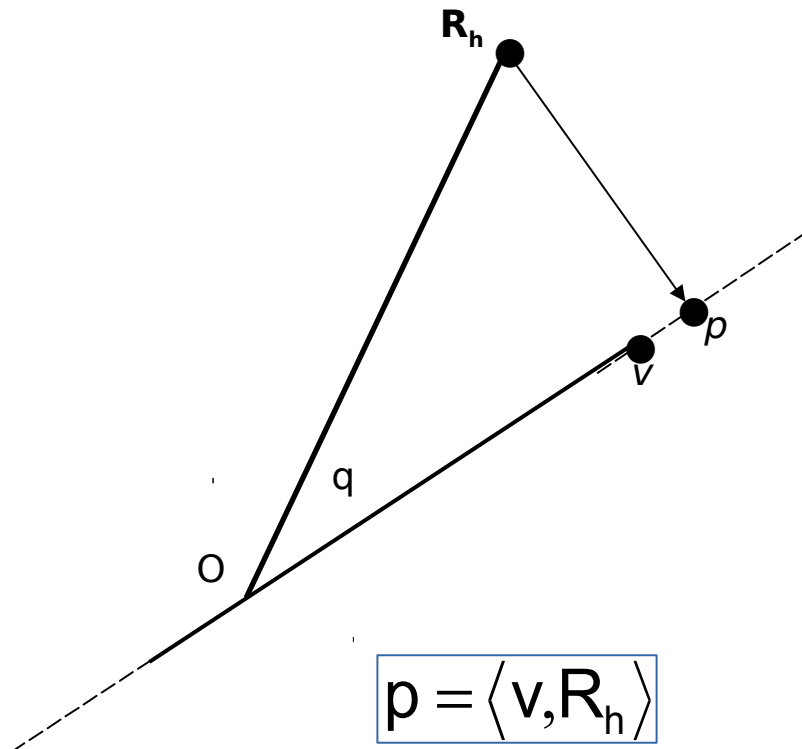
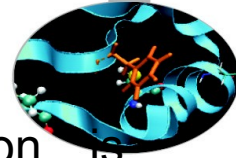


$$\text{RMSF} = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - \langle r \rangle)^2}$$



RMSF is a simple tool to measure the rigidity of the polypeptide chain. It calculates the deviations of C-alpha atoms coordinates from their average position. The flexibility pattern reflects the location of secondary structure elements in the protein structure.

PCA: how it works



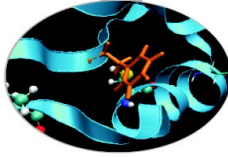
p is the projection of vector \mathbf{R}_h onto unit vector \mathbf{v} (dot product between \mathbf{v} and \mathbf{R}_h)

Let's assume our simulation is defined by the vector \mathbf{R}_n , that simply consists of the set of cartesian coordinates of Ca atoms at a given time step.

Question: what is the unity vector so that projection of \mathbf{R}_h on vector \mathbf{v} is the largest possible?

Answer: it is the vector \mathbf{v} , so that the variance of the projected point p of \mathbf{R}_n onto \mathbf{v} is the largest possible

Eigenvalue equation



The average of projected points onto \mathbf{v} is:

$$\mu(\mathbf{v}) = \langle \mathbf{v}, \bar{\mathbf{x}} \rangle$$

Variance of projected point onto \mathbf{v} is:

$$\sigma^2(\mathbf{v}) = \langle \mathbf{C}\mathbf{v}, \mathbf{v} \rangle$$

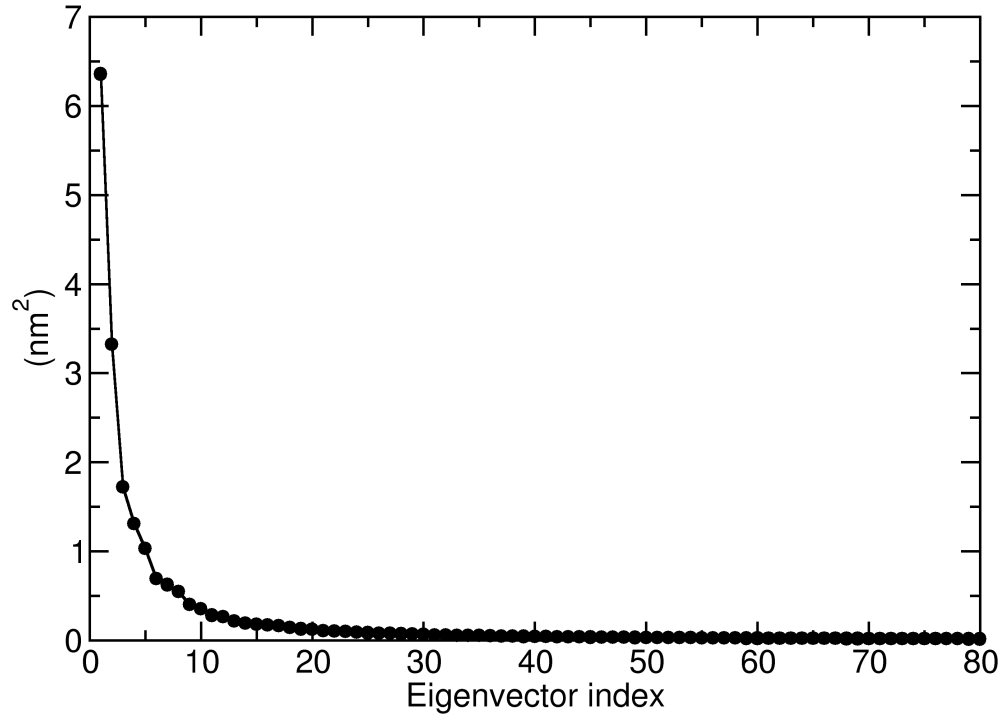
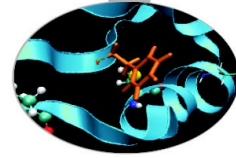
Variance of projected points along vector \mathbf{v} can be expressed in terms of dot product between \mathbf{v} and $\mathbf{C}\mathbf{v}$.

$$\begin{array}{ccc} \mathbf{C}\mathbf{v} = \lambda\mathbf{v} & & \\ \downarrow & & \uparrow \\ \text{Eigenvectors} & & \text{Eigenvalues} \end{array}$$

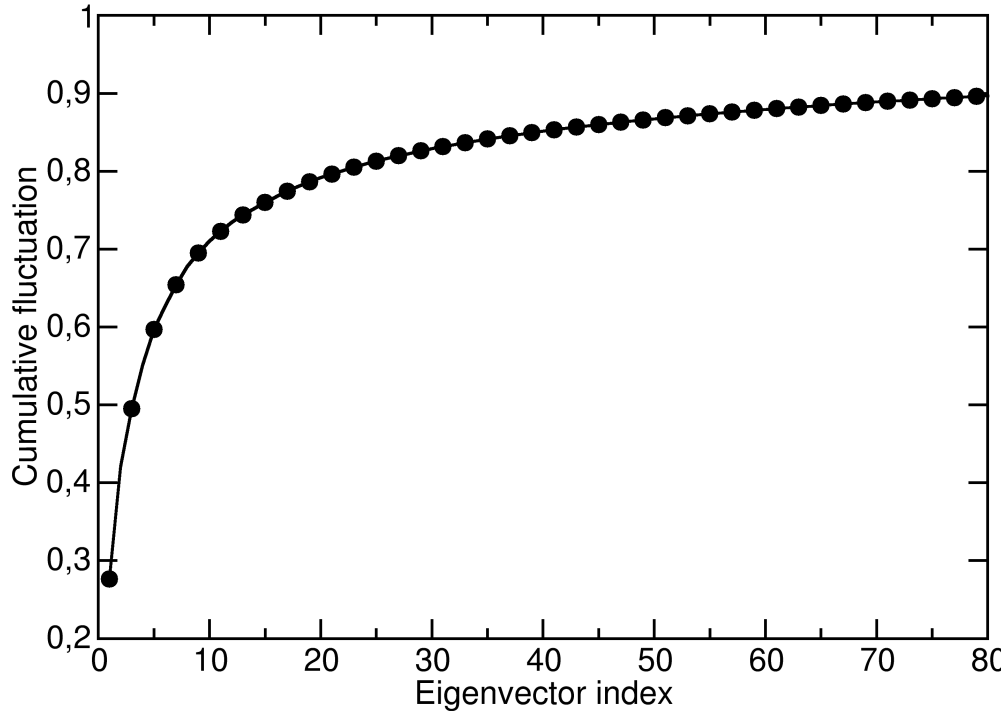
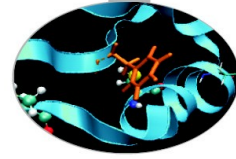
Eigenvectors represents direction where the σ^2 returns its maximum value.

Moreover, it can be shown that σ^2 are numerically equivalent to calculated eigenvalues

Essential Dynamics of Proteins



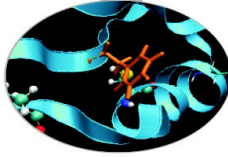
Eigenvalues are sorted in descending order: the first one corresponds to the maximum variance of the projected points. The corresponding eigenvectors are the best principal components of associated eigenvalues.



$$CF = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_n}{\sum_{i=1}^N \lambda_i}$$

The essential space, or subspace, of a biological protein is defined by the first 10 eigenvectors of the fluctuations covariance matrix. Indeed, it can be shown that about 70-75 % of all cumulative protein fluctuation is spanned by the first 10 principal components (eigenvectors)

Essential Dynamics: workflow in GROMACS

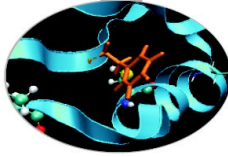


Least square fit of protein coordinates on respect to reference structure to remove roto-translation in the simulation box.

Calculate elements of the positional fluctuations covariance matrix of the $C\alpha$ protein carbon atoms.

Diagonalization of the covariance matrix and output of the corresponding eigenvectors and eigenvalues.

Sort eigenvector in descending eigenvalue index and determine principal components



Essential Dynamics Analysis is based on the computation of the elements of positional fluctuations covariance matrix of protein C α carbon atoms as follows:

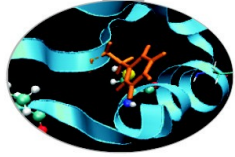
$$\Gamma_{ij} = \frac{1}{n} \sum_{h=1}^n (x_{hi} - \bar{x}_i) \times (x_{hj} - \bar{x}_j)$$

```
g_covar -f traj.xtc -s reference.gro -b start -e end -ascii
```

Output files:

Eigenvec.trr	→ eigenvector traj. file
Eigenval.xvg	→ eigenvalue set file
Covar.dat	→ covariance matrix in raw data format

Principal components analysis

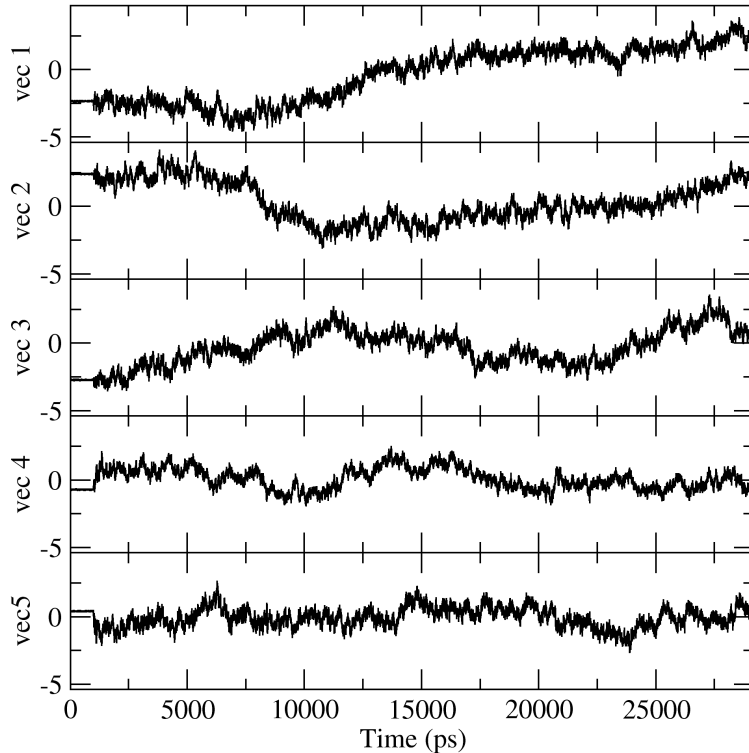
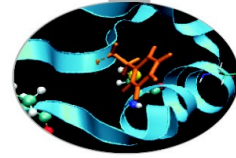


```
g_anaeig -f trajectory.xtc -v eigenvec.trr -eig eigenval.xvg  
-s reference.gro -b start -e end -first eig-first -last eig-last
```

g_anaeig reads a set of eigenvectors and eigenvalues as input files and returns a set of output files that can be selected using appropriate flags:
Here are some examples:

- proj to project an MD trajectory along a selected eigenvector
- rmsf to calculate the RMSF along a selected eigenvector
- extr to compute linear combinations of trajectory and selected eigenvectors
- filt to filter trajectory along selected eigenvector

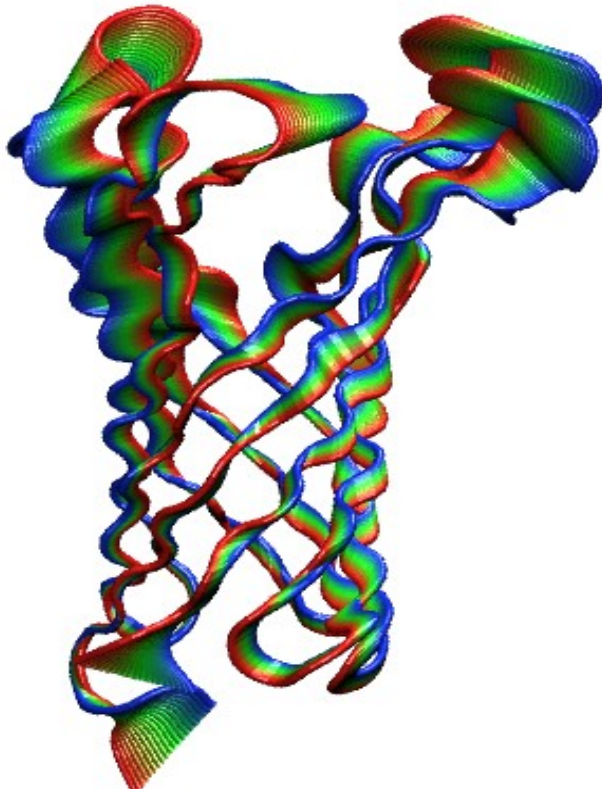
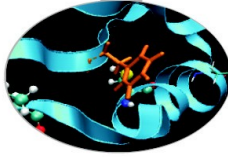
g_anaeig: output of flag -proj



By default, 8 eigenvectors are considered for output using g_anaeig. This option can be set by using the flags `-first` and `-end`

```
g_anaeig -f trajectory.xtc -v eigenvec.trr -eig eigenval.xvg -s reference.gro  
-proj proj.xvg -first 1 -last 5
```

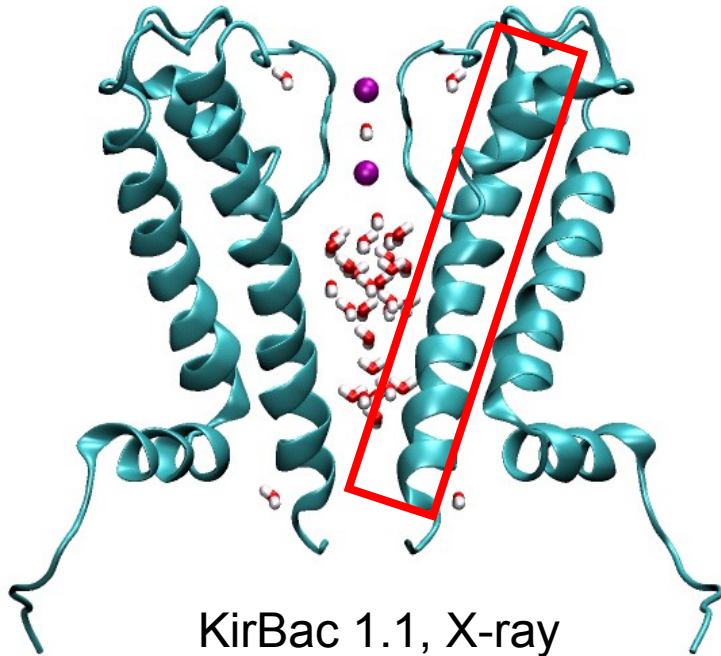
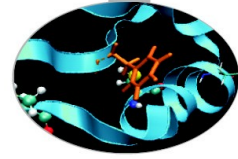
g_anaeig: flag -extr



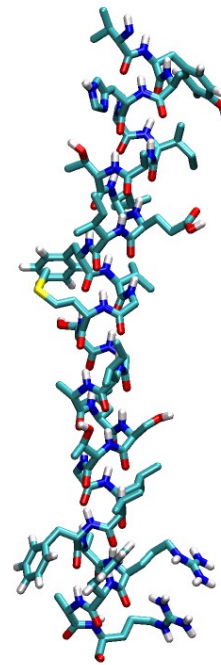
La dinamica essenziale ci aiuta a studiare i moti concertati tra gruppi di atomi all'interno di una struttura proteica. Nell'esempio si osserva un movimento concertato tra i loop extracellulari della porina OmpA lungo il primo autovettore.

```
g_anaeig -f trajectory.xtc -v eigenvec.trr -eig eigenval.xvg -s reference.gro  
-extr extreme.pdb -first 1 -last 3 -nframes 50
```

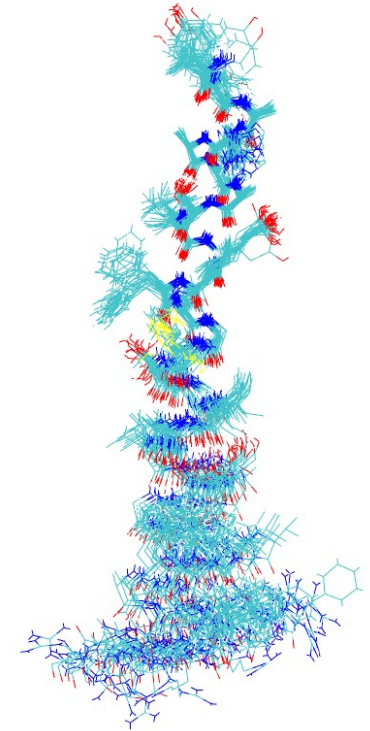
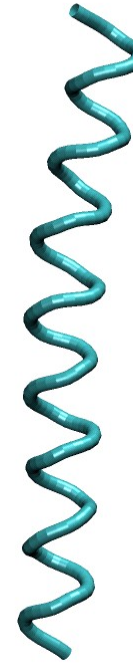
g_anaeig: the flag -filt



KirBac 1.1, X-ray

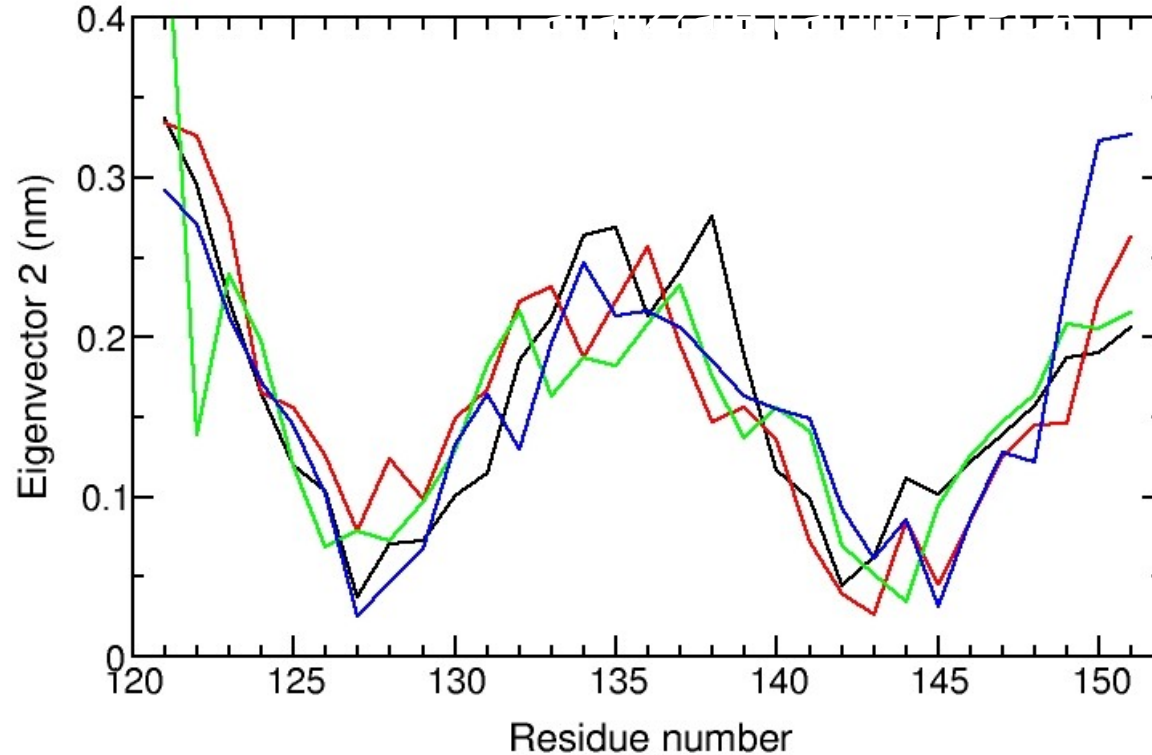
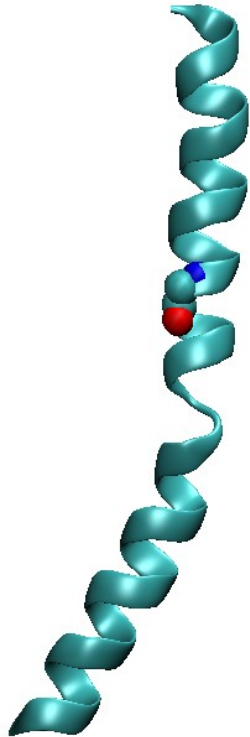
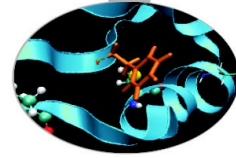


X-ray structure



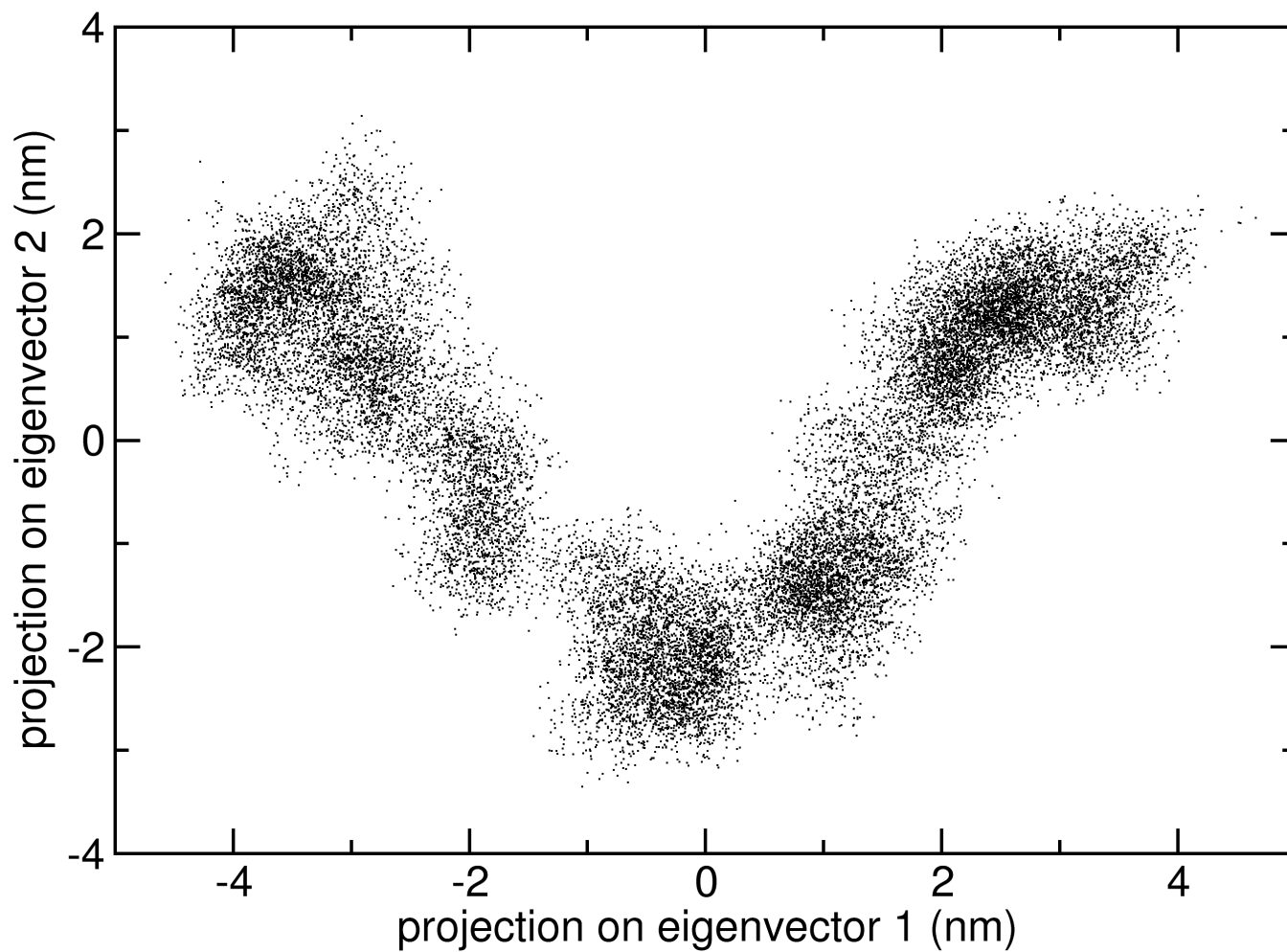
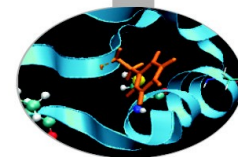
MD simulation

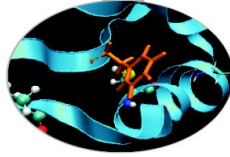
g_anaeig: the -rmsf flag



RMSF Analysis by means of principal components analysis. Fluctuations analysis by PCA on K channels simulations revealed that top flexibility on helix S6 is achieved at a conserved Gly-134 residue that is the most flexible in the polypeptide chain.

g_anaeig: the -2d or -3d flag





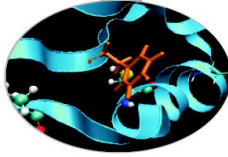
Input file on PLX:

```
/gpfs/scratch/userinternal/agrottes/Corso-PATC-2014/Tutorial3/Essential-Dynamics
```

Input file on Eurora:

```
/gpfs/scratch/userinternal/agrottes/Corso-PATC/Tutorial3/Essential-Dynamics
```

- `cp -r <source_dir> ./`
- `cd Essential-Dynamics`
- Run `g_covar` on file `total19ns.xtc` using `start_prot.gro` as reference
- Run `g_anaeig` with option `-proj -extr -filt` and `-2d`
- Get the first principal plane
- Run `g_rms` and `g_rmsf`



To analyze results on PLX:

- Module load autoload vmd
- Module load xmgrace

Alternatively, download RCM on:

<http://www.hpc.cineca.it/content/remote-visualization-rcm#download>

