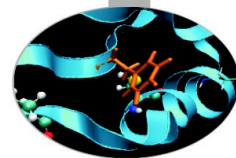# Analysis of MD trajectories (Essential Dynamics of Proteins)

Neva Bešker
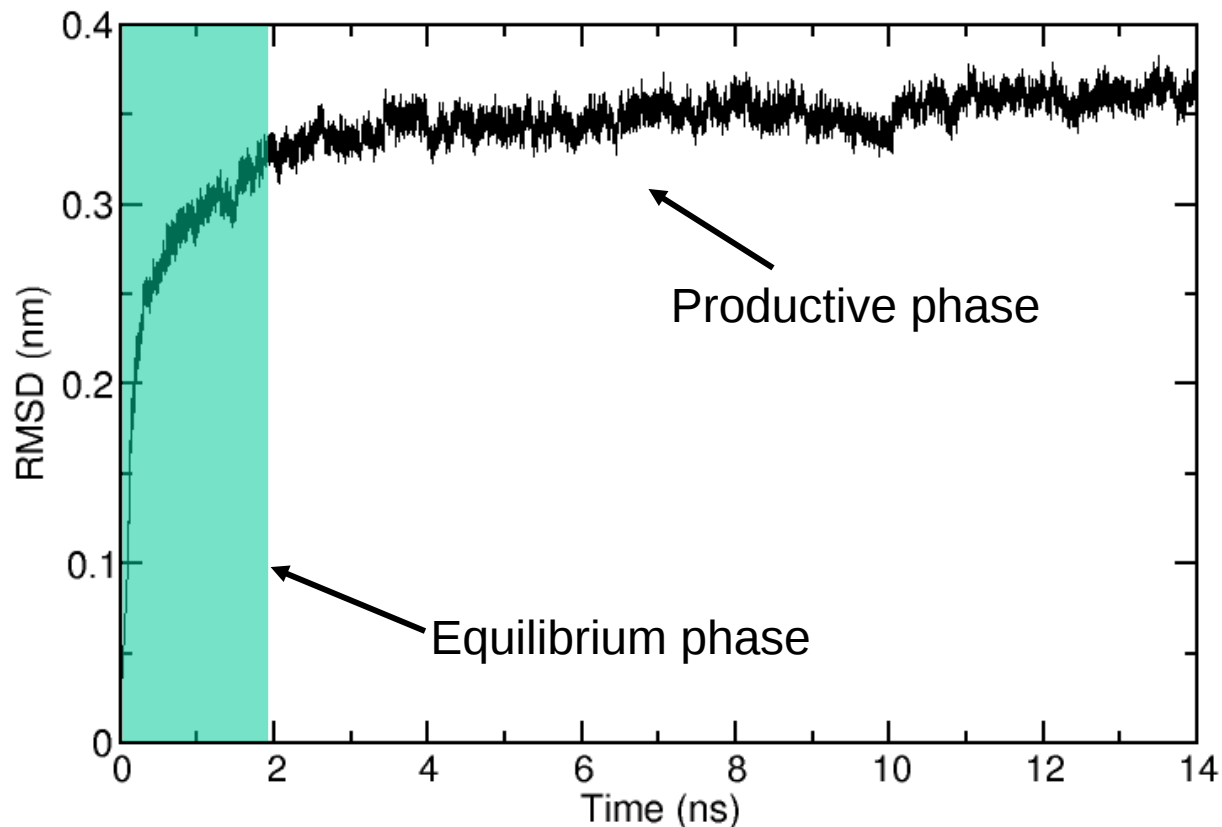
Did we reach equilibrium…?
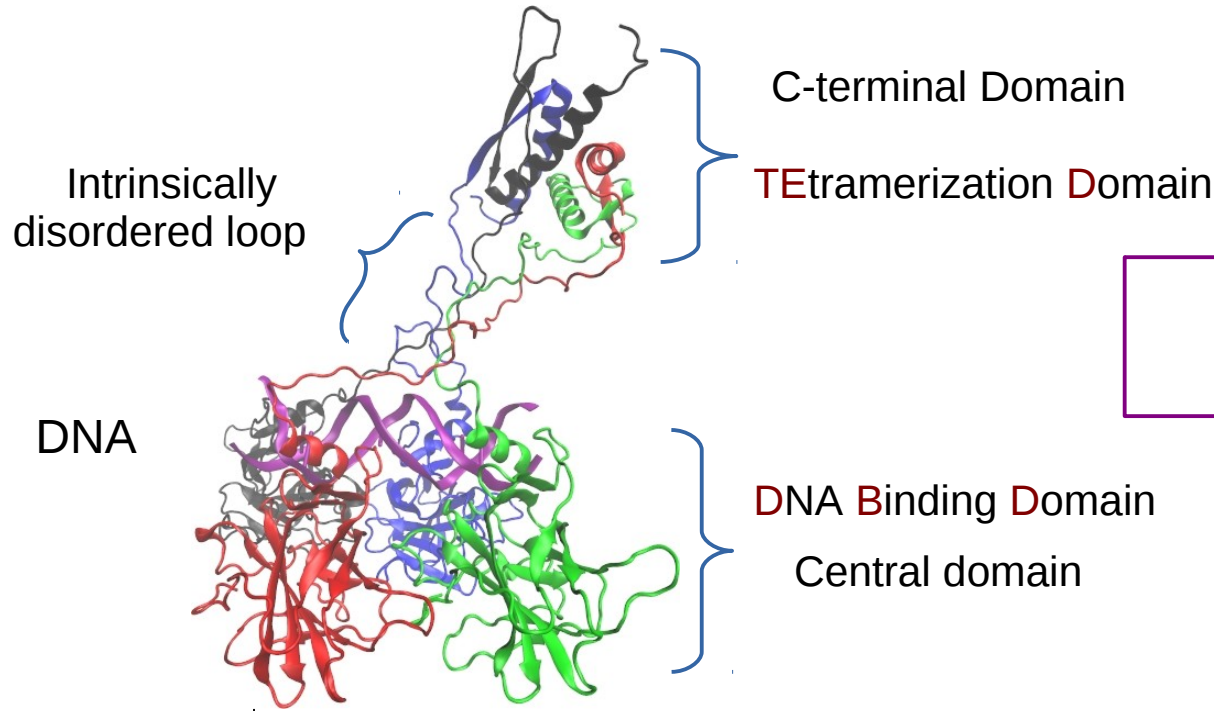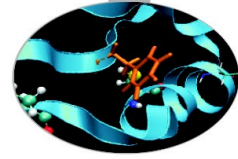
$$= \left[\frac{1}{M} \sum_{i=1}^{N} m_i \| \mathbf{r}_i(t_1) - \mathbf{r}_i(t_2) \|^2 \right]^{\frac{1}{2}}$$
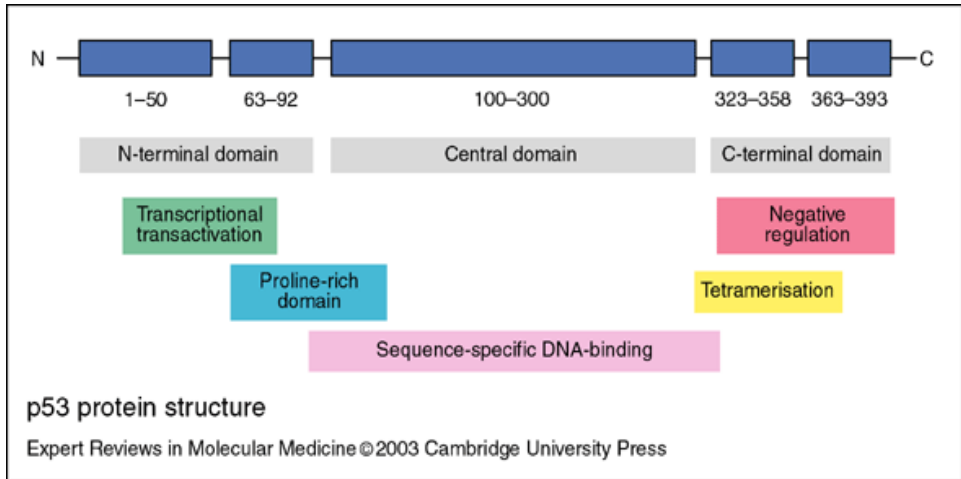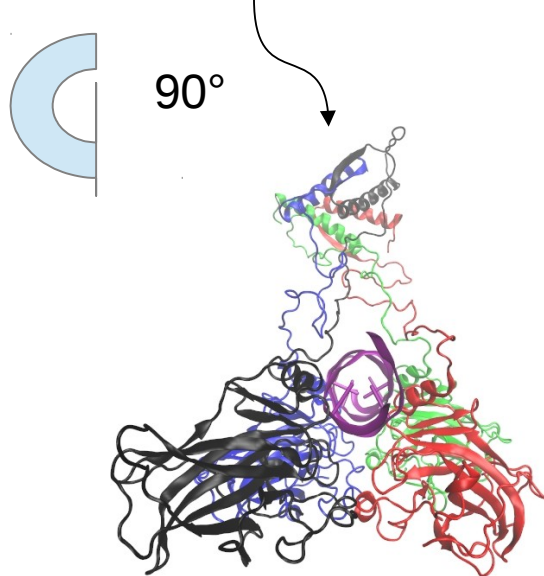
Gromacs tool: gmx rms

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 Cα carbon atoms position with respect to start.
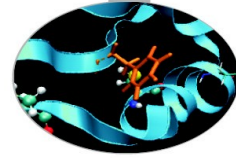

Productive phase

Equilibrium phase

p53 example

C-terminal Domain

TEtramerization Domain

Intrinsically disordered loop

4 MD of 25 ns
~ 347000 atoms

DNA

DNA Binding Domain

Central domain

90°

N | 1–50 | 63–92 | 100–300 | 323–358 | 363–393 | C

N-terminal domain | Central domain | C-terminal domain

Transcriptional transactivation

Negative regulation

Proline-rich domain

Tetramerisation

Sequence-specific DNA-binding

p53 protein structure

Expert Reviews in Molecular Medicine © 2003 Cambridge University Press

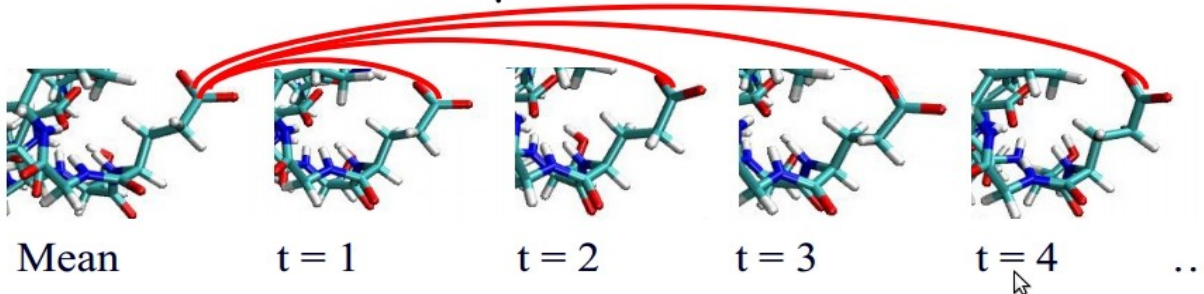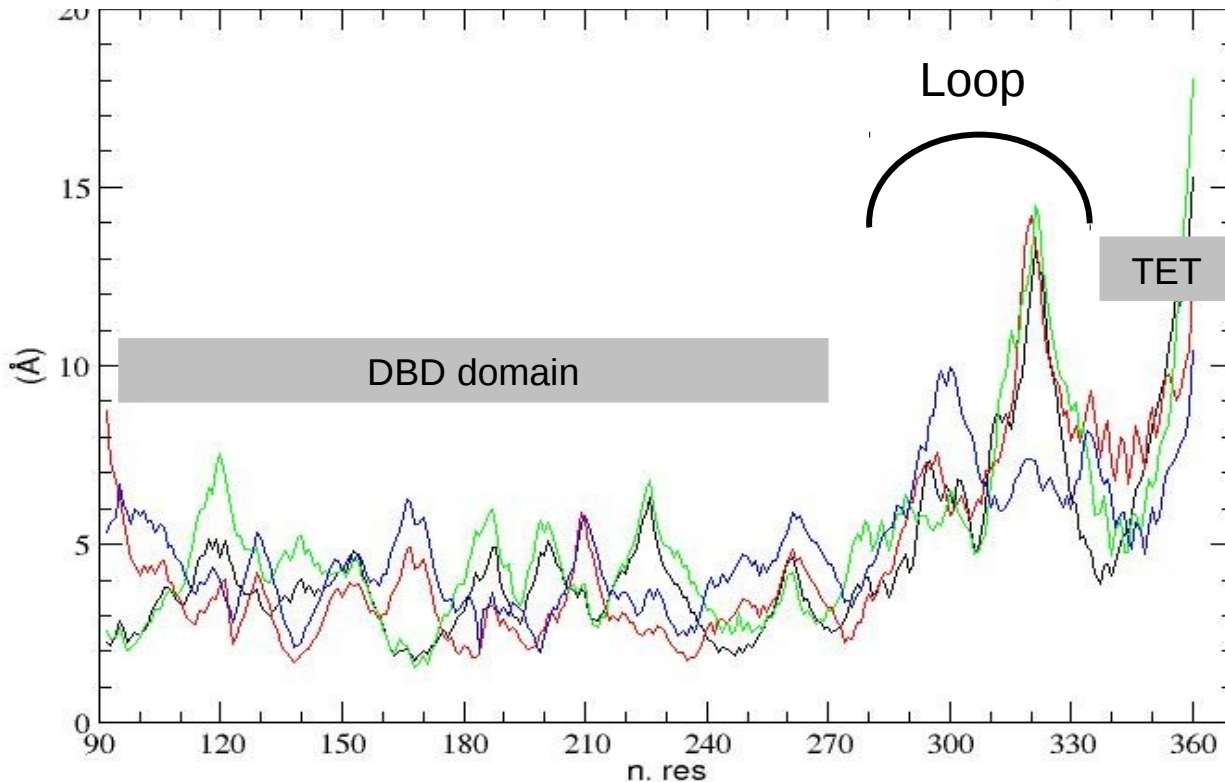# Measuring chain flexibility

p53 in water
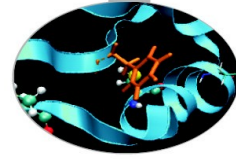
$$RMSF_{(v)} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2}$$

$T = \#(Frames)$

RMSF is an analysis to measure the flexibility of the polypeptide chain. It calculates the fluctuation of C-alpha atoms coordinates from their average position.

D'Abramo et al. Oncogene, October 2015;doi: 10.1038/onc.2015.388

Essential Dynamics analysis

Essential dynamics analysis

How represent the "principal motion directions" of large systems?
The essential dynamics (ED)* is a technique able to represent the principal motion directions by a set of eigenvectors
(look at eigenvectors as important motion directions!)
Example - reducing bi-dimensional to monodimensional
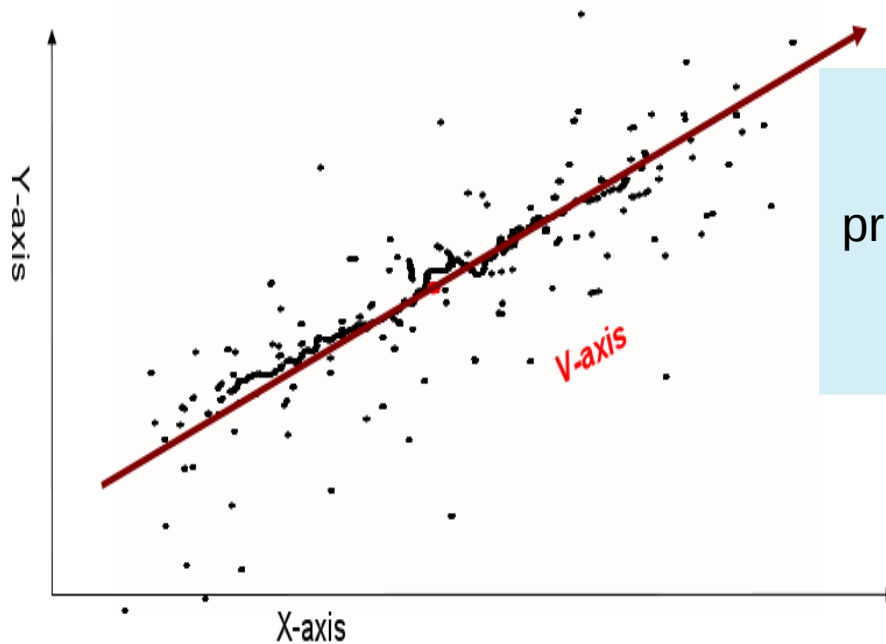
used for biological systems

principal modes ⟷ biological function

Based on PCA

Essential Dynamics: workflow in GROMACS
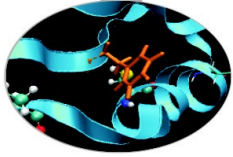
Least square fit of protein coordinates with respect to a reference structure to remove roto-translation in the simulation box.
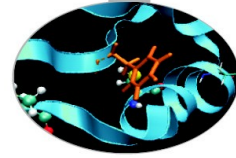
Calculate elements of the positional fluctuations covariance matrix of the Cα protein carbon atoms.

$$C_{ij} = \left\langle M_{ii}^{\frac{1}{2}}(x_i - \langle x_i \rangle) M_{jj}^{\frac{1}{2}}(x_j - \langle x_j \rangle) \right\rangle$$

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

Sort eigenvector in descending eigenvalue index and determine principal components

Eigenvalue equation

$$Cv = \lambda v$$

Eigenvectors   Eigenvalues

Eigenvectors represent directions where the σ^2 returns its maximum value.

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

Eigenvectors, also called principal or essential modes.

R defines a transformation to a new coordinate system. The trajectory can be projected on the principal modes to give the principal components
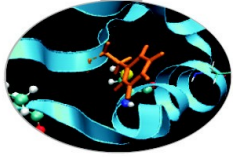
Essential Dynamics of Proteins
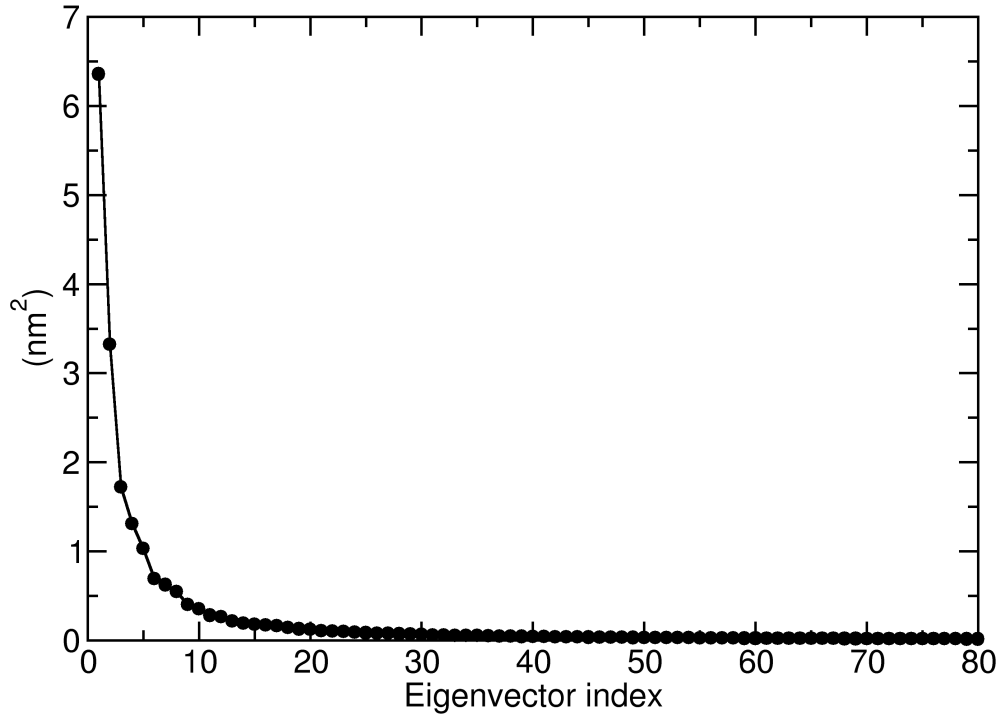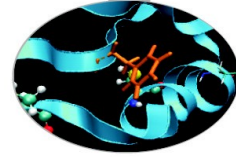
Eigenvalues are sorted in descending order: the first one corresponds to the maximum variance of the projected points. The corresponding eigenvectors are the best principal components of associated eigenvalues.

Essential Dynamics of Proteins
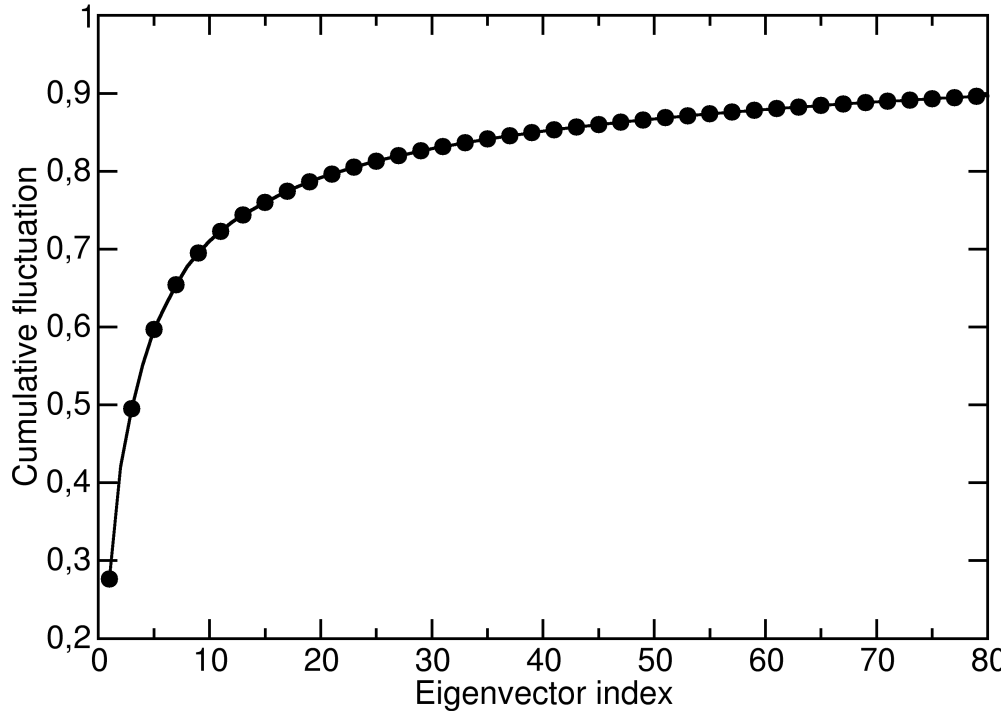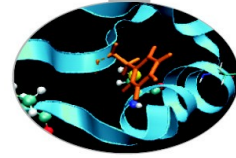
$$CF = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_n}{\sum\limits_{i=1}^{N} \lambda_i}$$

It can be shown that about 70-75 % of all cumulative protein fluctuation is spanned by the first 10 principal components (eigenvectors)

Essential Dynamics Analysis is based on the computation of the elements of positional fluctuations covariance matrix of protein Cα carbon atoms as follows:
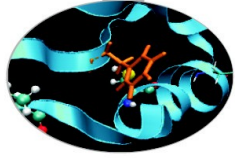
$$\Gamma_{ij} = \frac{1}{n}\sum_{h=1}^{n}\left(x_{hi} - \overline{x}_{i}\right) \times \left(x_{hj} - \overline{x}_{j}\right)$$

gmx covar –f traj.xtc –s reference.tpr –b start –e end –ascii

Output files:

|	
Eigenvec.trr	→ eigenvectors file
Eigenval.xvg	→ eigenvalues file
Covar.dat	→ covariance matrix in ascii

Principal components analysis

gmx anaeig –f trajectory.xtc –v eigenvec.trr –eig eigenval.xvg
        –s reference.tpr –b *start* –e *end* -first *eig-first* -last *eig-last*

gmx 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
-2d   to project an MD trajectory in two dimensions of selected eigenvectors
      (essential subspace)
-rmsf   to plot the RMS fluctuation per atom of selected eigenvectors
-comp   to extract the eigenvector (atomic) components
-extr   to extract structures projecting on the extreme of the selected eigenvectors
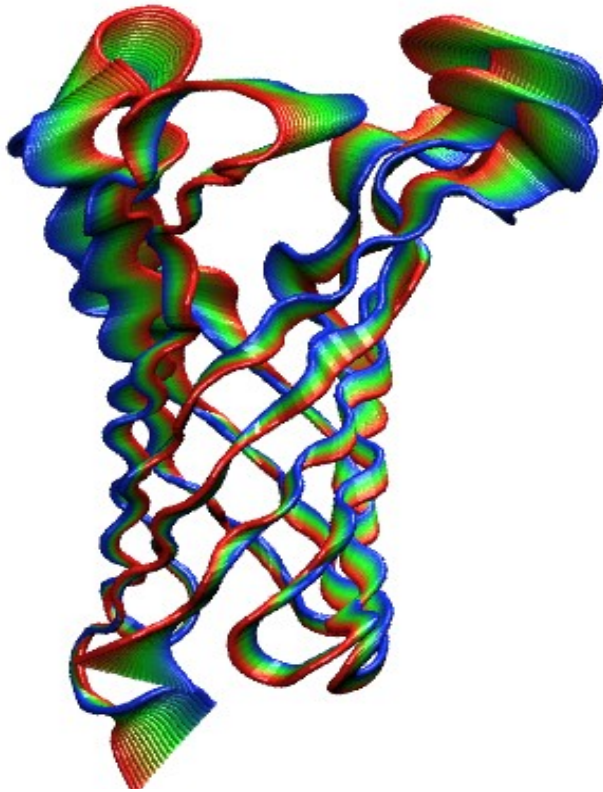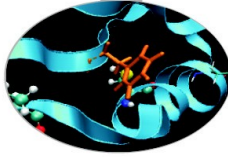-filt   to filter trajectory along selected eigenvectors

gmx anaeig: output of flag –proj
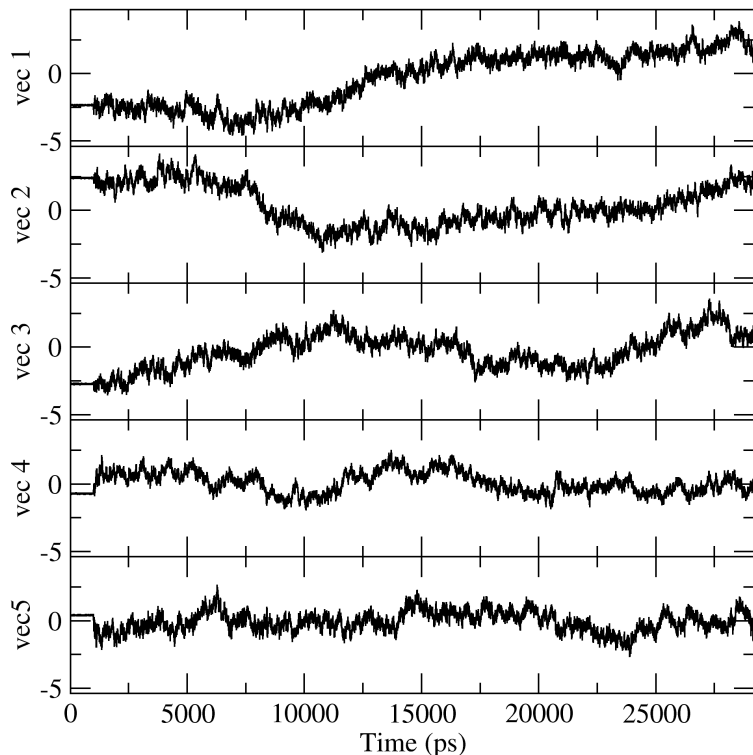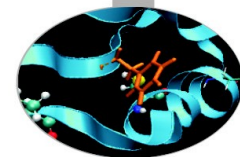
8 eigenvectors are considered for output using gmx anaeig. This option can be set by using the flags `-first` and `-last`
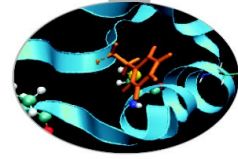
gmx anaeig -f trajectory.xtc -v eigenvec.trr -eig eigenval.xvg -s reference.tpr -proj proj.xvg -first 1 -last 5
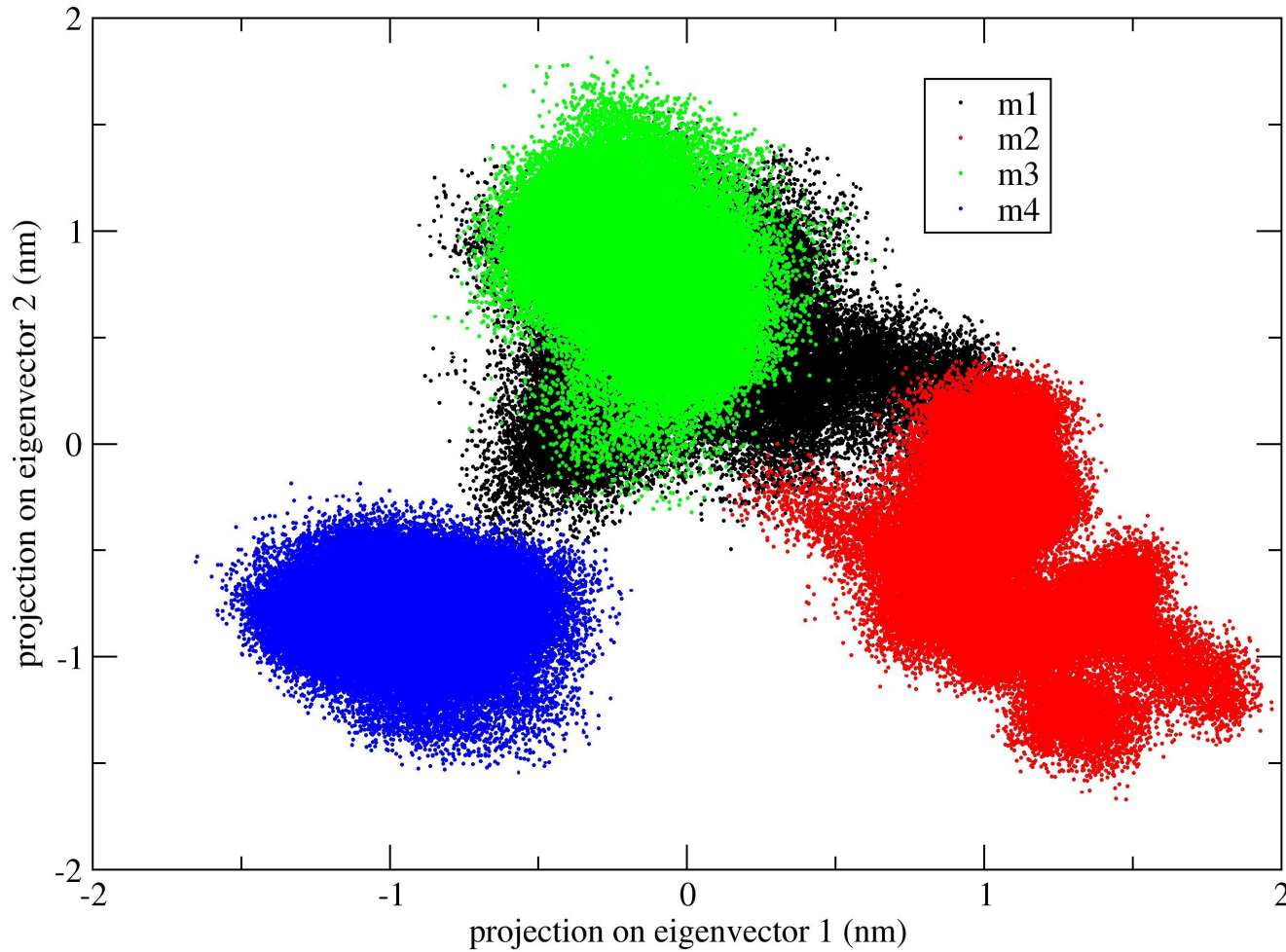
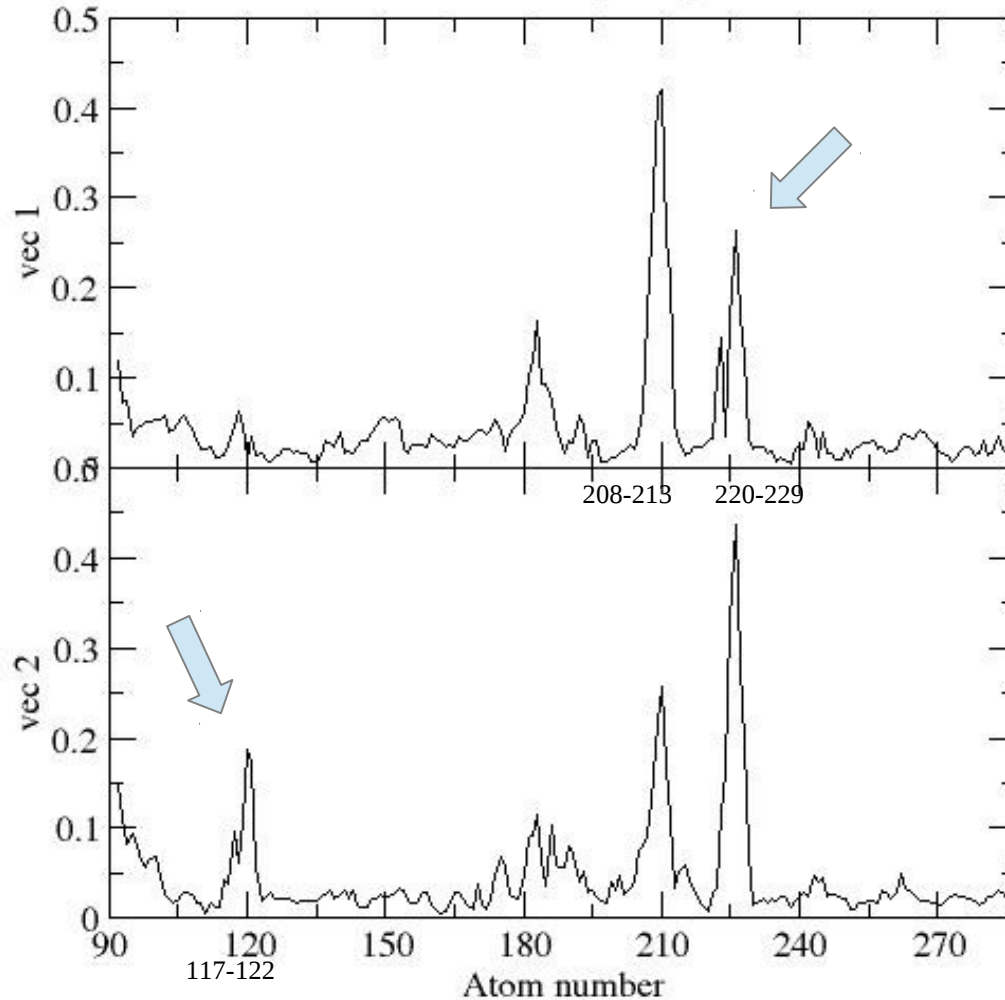The concerted movements along the first eigenvectors of the extracellular loops of the OmpA porine are shown
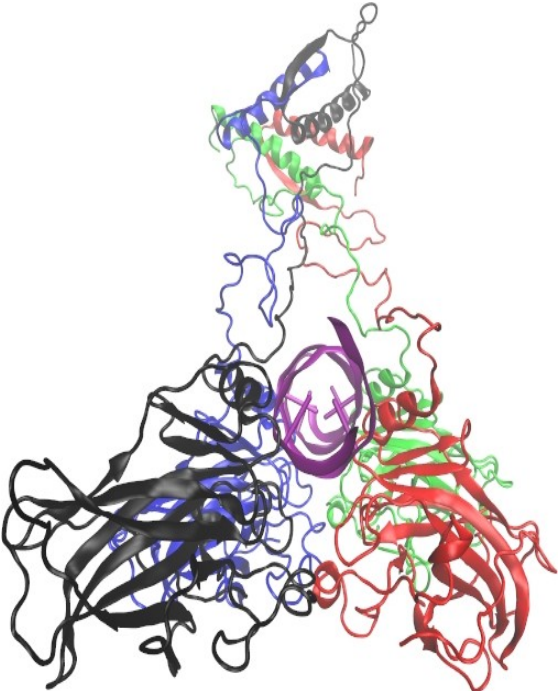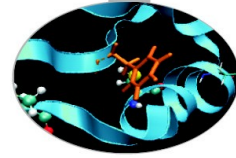
gmx anaeig -f trajectory.xtc -v eigenvec.trr -eig eigenval.xvg -s reference.gro -extr extreme.pdb -first 1 -last 3 -nframes 50
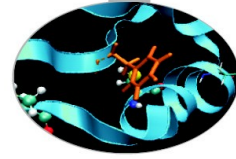
gmx anaeig: the -2d flag

p53 example

DNA structure

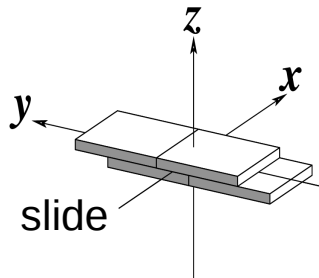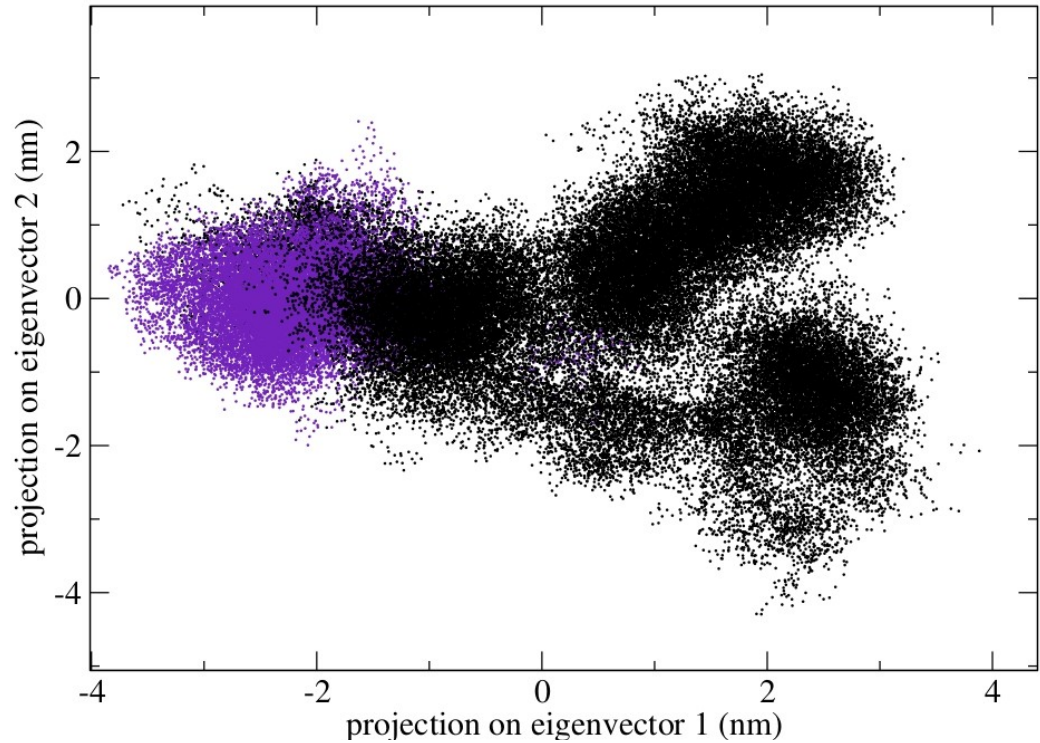DNA is in the B form

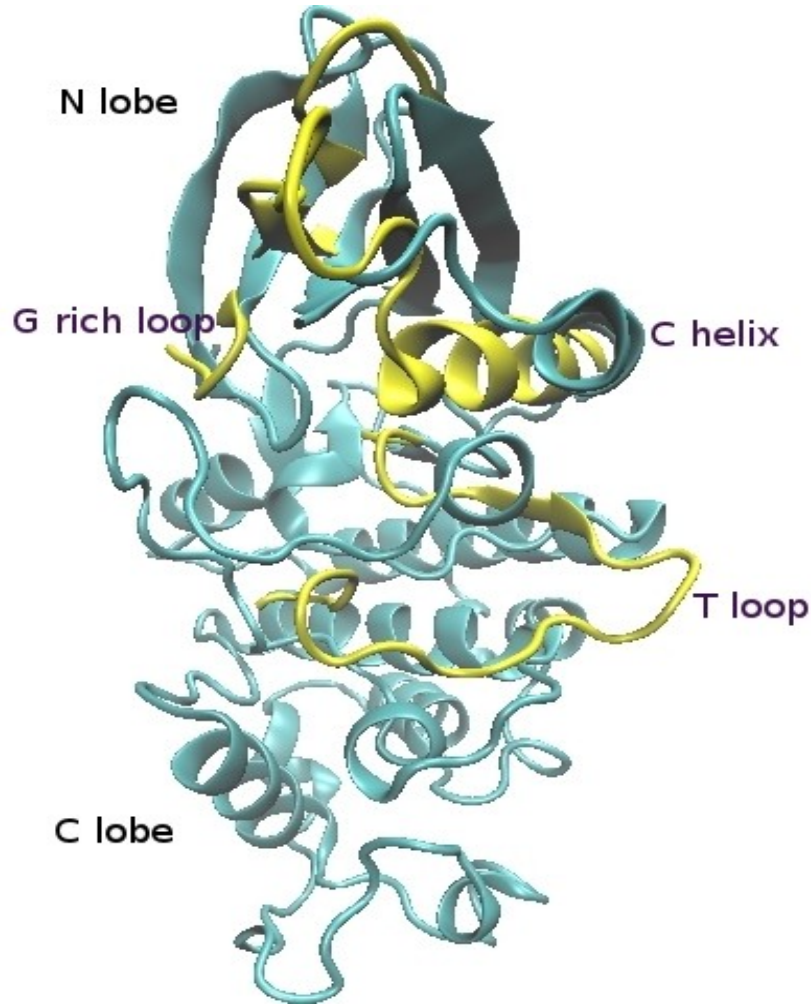The conformations in which both the DNA roll and twist parameters for the central TA bases are significantly different from their corresponding values found in standard B-DNA are shown in purple. The essential subspace in which the DNA is distorted is separated along the first eigenvector from the great majority of the p53 tetramer conformations visited during the simulation.

twist

roll

slide

CDK2 example

N lobe

G rich loop

C helix

C lobe

T loop

Cyclin A

Cyclin A

partially active activation loop open

P

Inactive form

activation loop closed

Different orientation of Cα helix

fully active
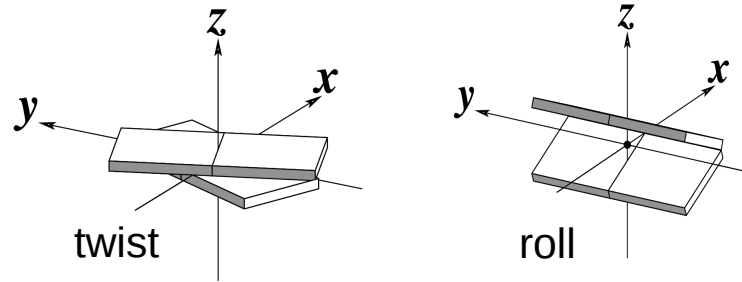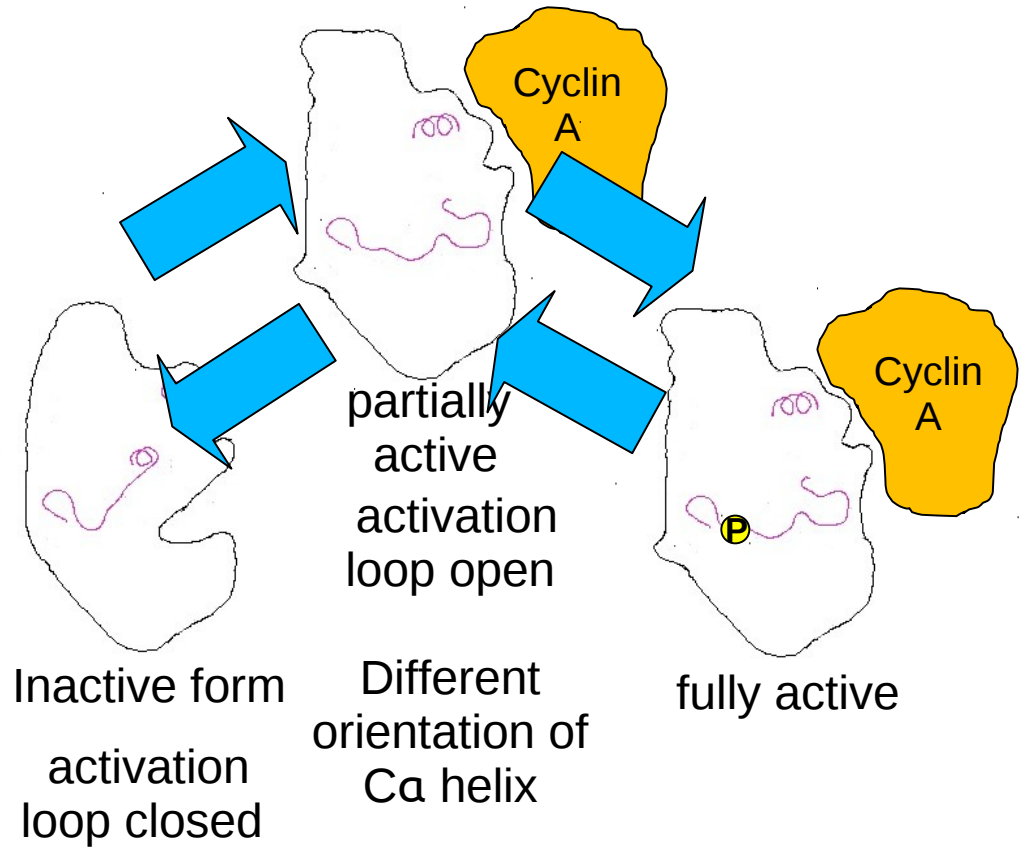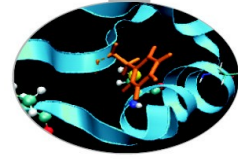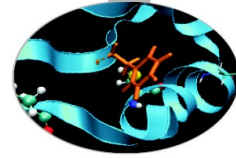
Essential Dynamics alghoritm to model conformational transition

Limit of the classic MD simulations: no transition in two MD simulations lasting 1 μs
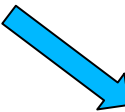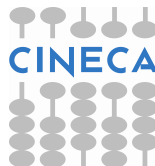
ED sampling

- is based on previous essential dynamics analysis of native protein movements in its stable conformational states. The system is constrained to move along variables defined by the unbiased MD simulations i.e. in the space of the selected eigenvectors
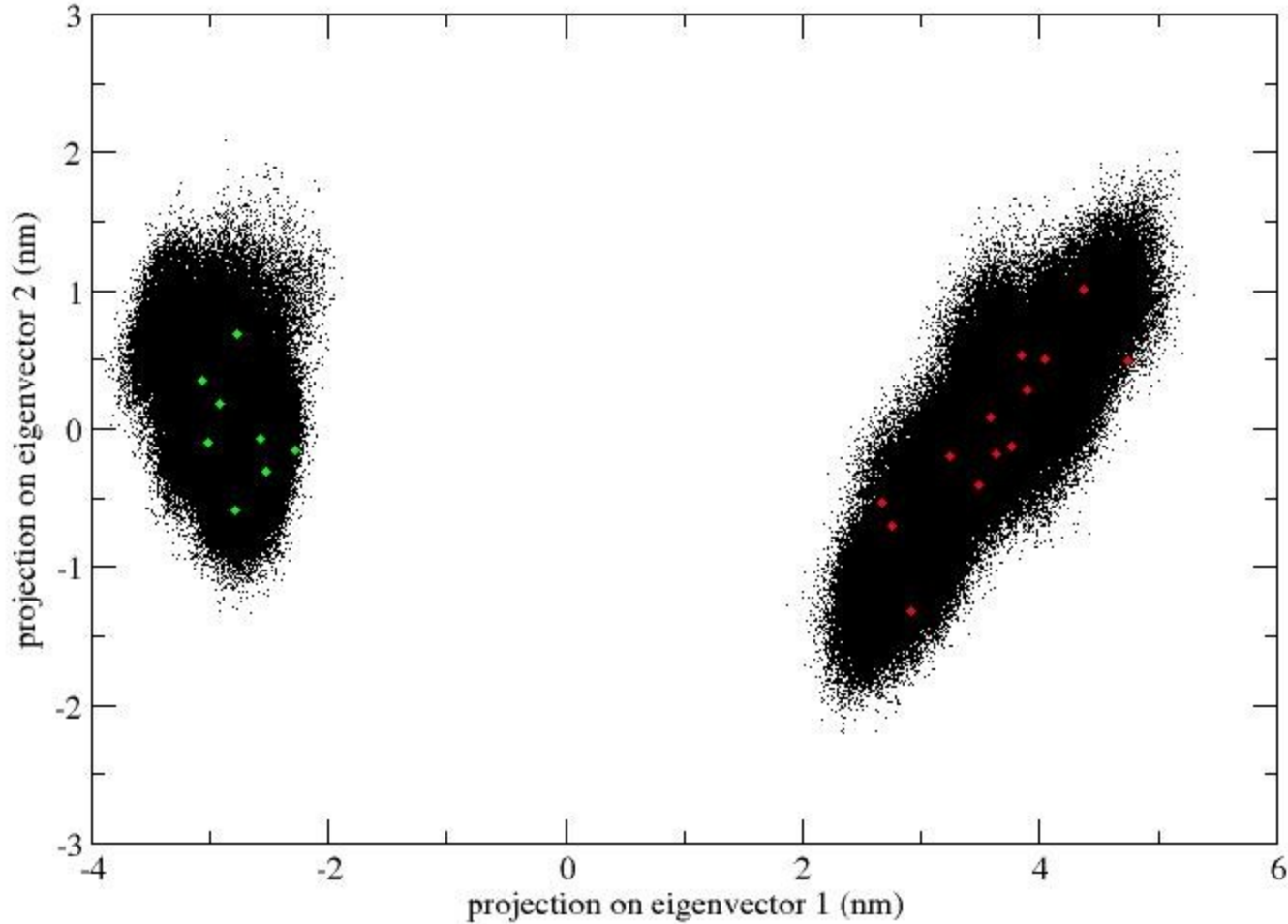
expansion procedure
increase the distance from a reference structure

contraction procedure
decrease the distance from a reference structure

$\Delta\xi = \Delta\xi_d$ per Δt Per $r \leq r_0$

$\Delta\xi = \Delta\xi_d + \Delta\xi_c$ per Δt Per $r > r_0$

Amadei et al., J. Biomol. Struct&Dynamics, 13, 4, (1996), 615-625
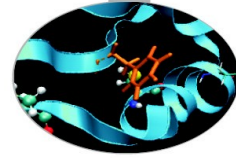
the two trajectories were concatenated, covariance matrices of positional fluctuations (C-alpha only) were built and diagonalized.

which eigenvector – obtained by means of principal component analysis – well discriminate between the two protein conformational states ?

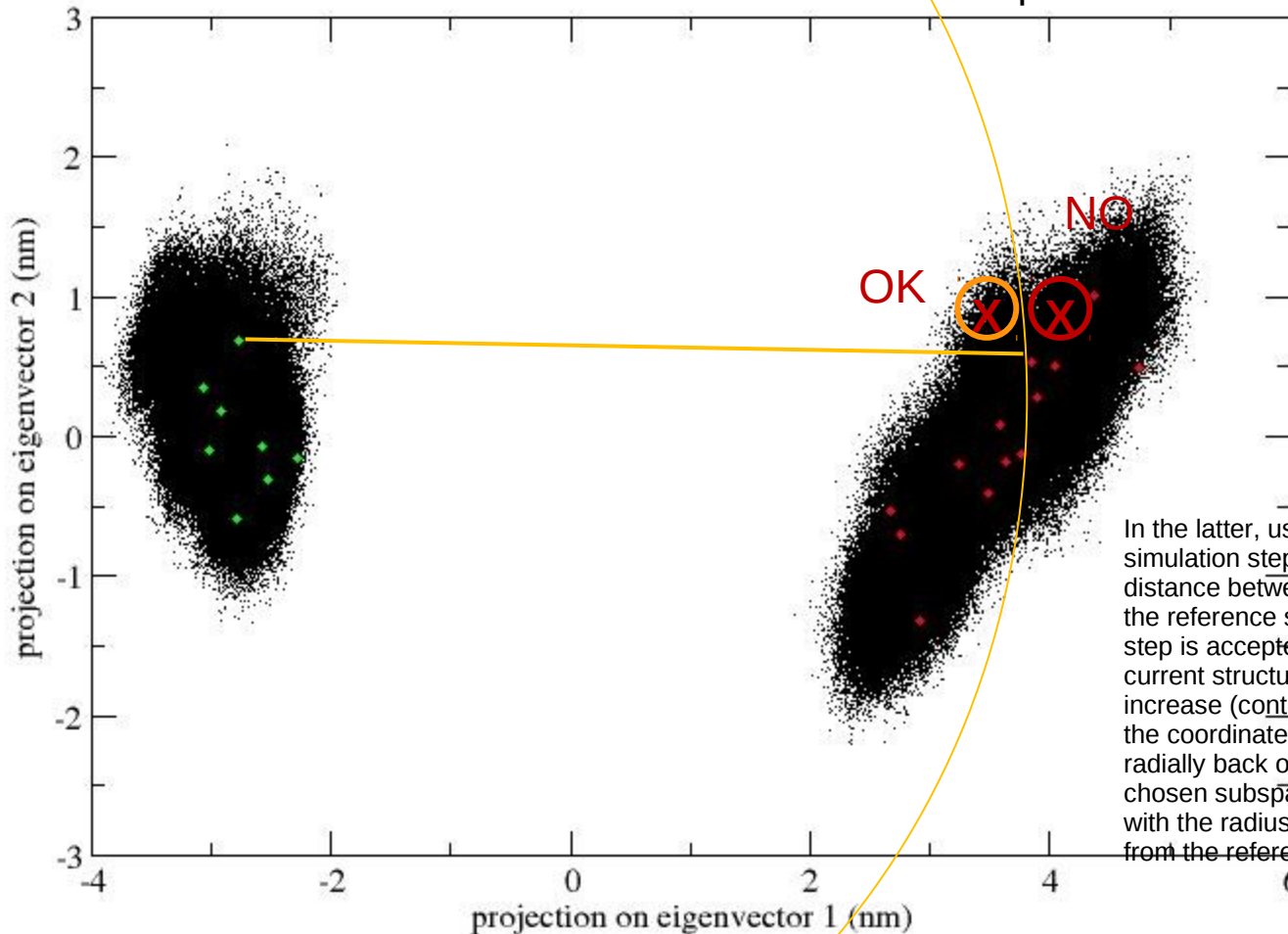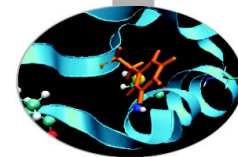Hypersphere with r=the distance between two structures on the essential subspace

In the latter, used in this work, a regular MD simulation step is performed and the distance between the current structure and the reference structure is calculated. The step is accepted if the distance between the current structure and the reference does not increase (contraction procedure); otherwise, the coordinates and velocities are corrected radially back onto the hypersphere (in the chosen subspace) centred in the reference, with the radius length given by the distance from the reference in the previous step.

THE SYSTEM IS CONSTRAINED TO MOVE TOWARDS A REFERENCE POINT (the origin of the hypersphere) IN THE ESSENTIAL SPACE
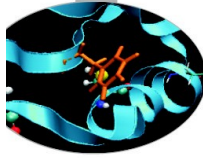
2D projection of trajectory
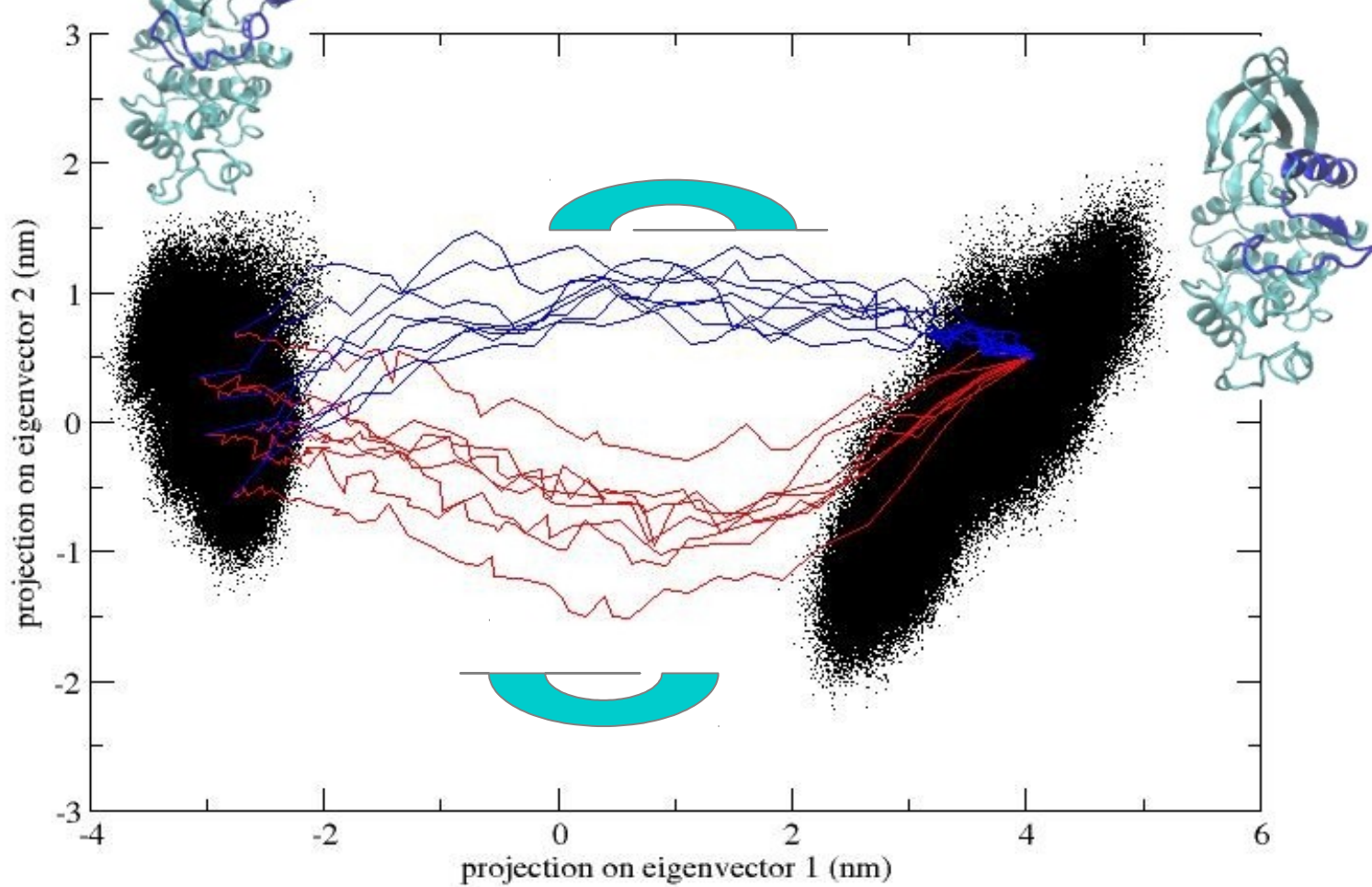
RMSD from the refence structure; conservation of the secondary structure, conservation of the H bonds
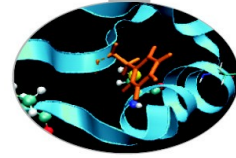
Besker et al J. Biomol Struct Dyn. 2014 Dec;32(12):1929-35.

Input file on Eurora:

/gpfs/scratch/userinternal/nbesker1/MDcourse-November2015.tar.gz

Copy Tutorial4.tar.gz file from Eurora:

- scp username@login.eurora.cineca.it

- Run gmx covar on file file.xtc using topol.tpr as reference in cdk2-closed

  example:

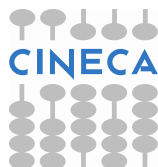 gmx covar -f cdk2-closed/traj-cdk2closed-mdcourse.xtc -s cdk2-closed/topol.tpr

  (select the group of interest, for example Calpha atoms)
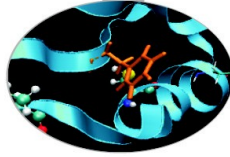
- Run gmx anaeig with option -proj -extr -filt and -2d

  example: gmx anaeig -f cdkd2-closed/traj-cdk2closed-mdcourse.xtc -v eigenvec.trr

   -first 1 -last 2 -s topol.tpr -2d

- Get the first principal plane

**Essential Dynamics Tutorial:
ED sampling of
open and closed CDK2**

- Concatenate the trajectories:
gmx trjcat -f ../cdk2-closed/dry.xtc ../cdk2-open/traj-cdk2open-mdcourse.xtc -o combined.xtc -settime
(digit 0 and then c)

- Calculate the covariance matrix
gmx covar -f combined.xtc -s ../cdk2-closed/topol.tpr
(select the group of interest for example Calpha carbons)

- Find the essential plane
gmx anaeig -f combined.xtc -v eigenvec.trr -2d -s cdk2-closed/topol.tpr -first 1 -last 2
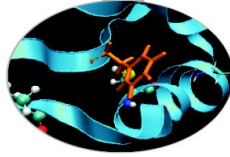
- You can extract the origin and a target structure for ed sampling or you can use files in dir Scripts_and_inputfiles/

- The control= projection of the extracted structures on the essential plane
gmx anaeig -f struct.gro -v eigenvec.trr -2d struct - 2dproj.xvg -first 1 -last 2 -s reference.gro
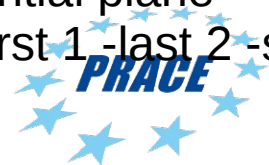
Essential Dynamics Tutorial:
ED sampling of
open and closed CDK2

 You can visualize plot of 2dproj.xvg and the structures or eigenval.xvg with gnuplot or xmgrace on local pc (no on eurora)

 scp username@login.eurora.cineca.it:/gpfs/scratch/usertrain/username/MDCorsi-November2015/2dproj .

• generates an essential dynamics (ED) sampling input file on eurora
gmx make_edi -radccon 1-298 -tar name-file.gro -o sam.edi
•

• Prepare tpr input file for the eds sampling
 gmx grompp -c Scripts_and_inputfiles/conf10ps.gro -p cdk2-closed/topol-closed.top -f run.mdp

• We have files tpr, sam.edi, and the script for the eds sampling
 qsub Scripts_and_inputfiles/eds.sh

• Look your output files and then project the trajectory of the transition on the essential subspace
 gmx anaeig ....