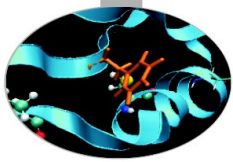
CINECA SCAI
SuperComputing Applications and Innovation

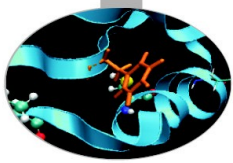


Analysis of MD trajectories (Essential Dynamics of Proteins)

Neva Bešker
Alessandro Grottesi

PRACE CINECA

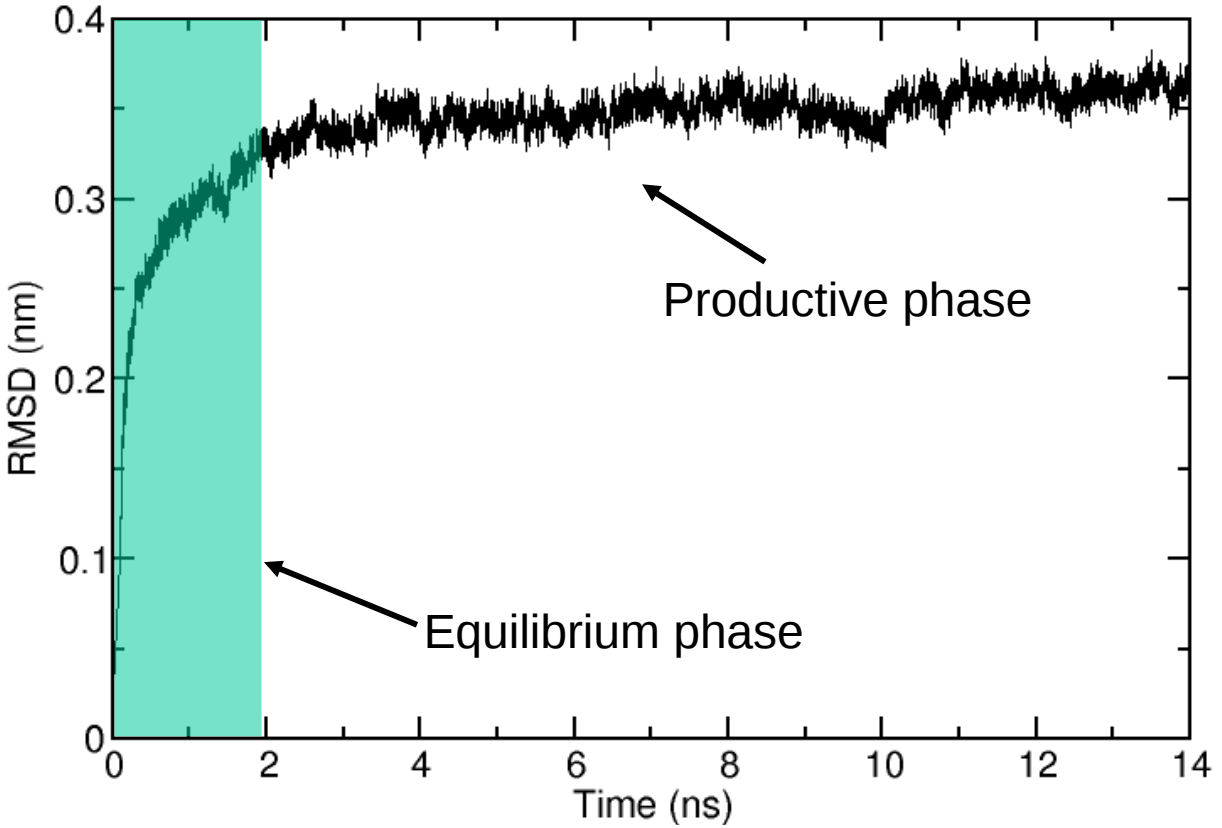
Did we reach equilibrium…?



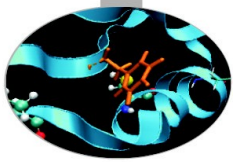
$$= \left[\frac{1}{M} \sum_{i=1}^{N} m_i \|\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)\|^2 \right]^{\frac{1}{2}}$$

Gromacs tool: g_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.



p53 example



loop intrinsecamente disordinato

Dominio C-terminale

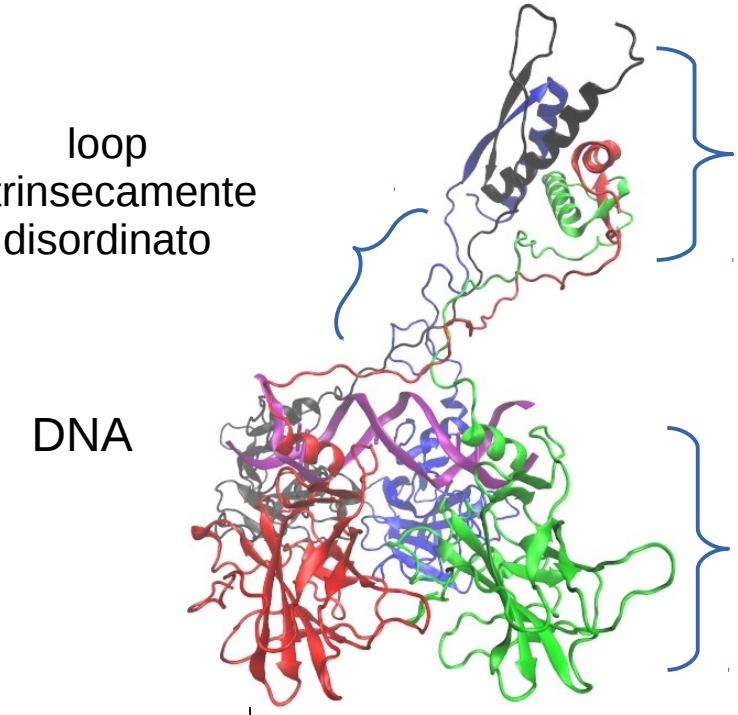
TEtramerization Domain

4MD da 25 ns
~ 347000 atomi

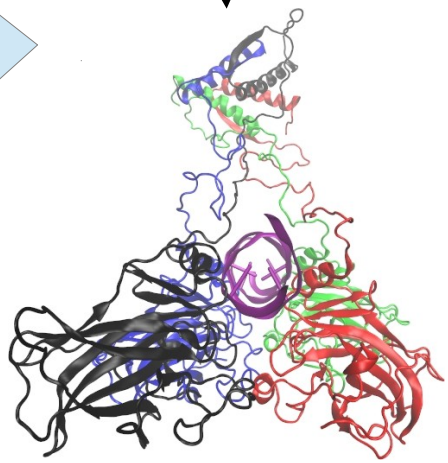
DNA

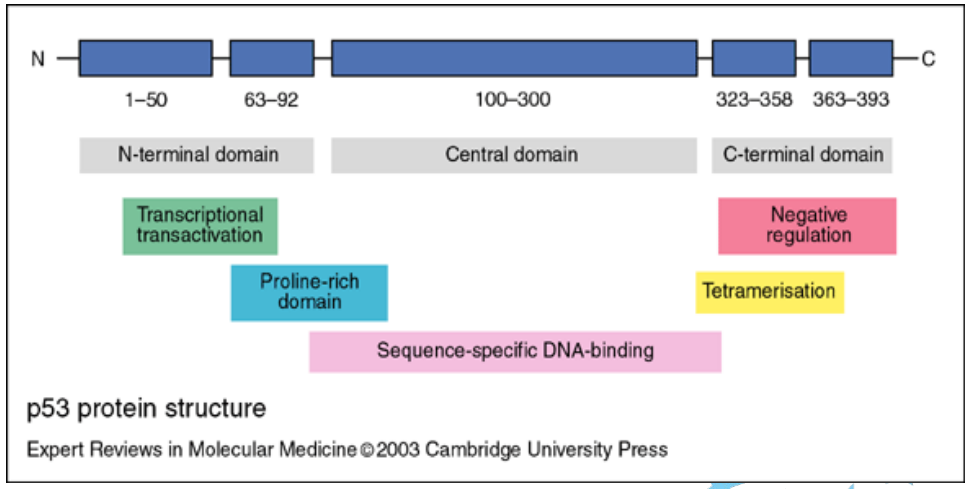
DNA Binding Domain

Dominio centrale



90°



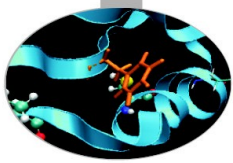


p53 protein structure

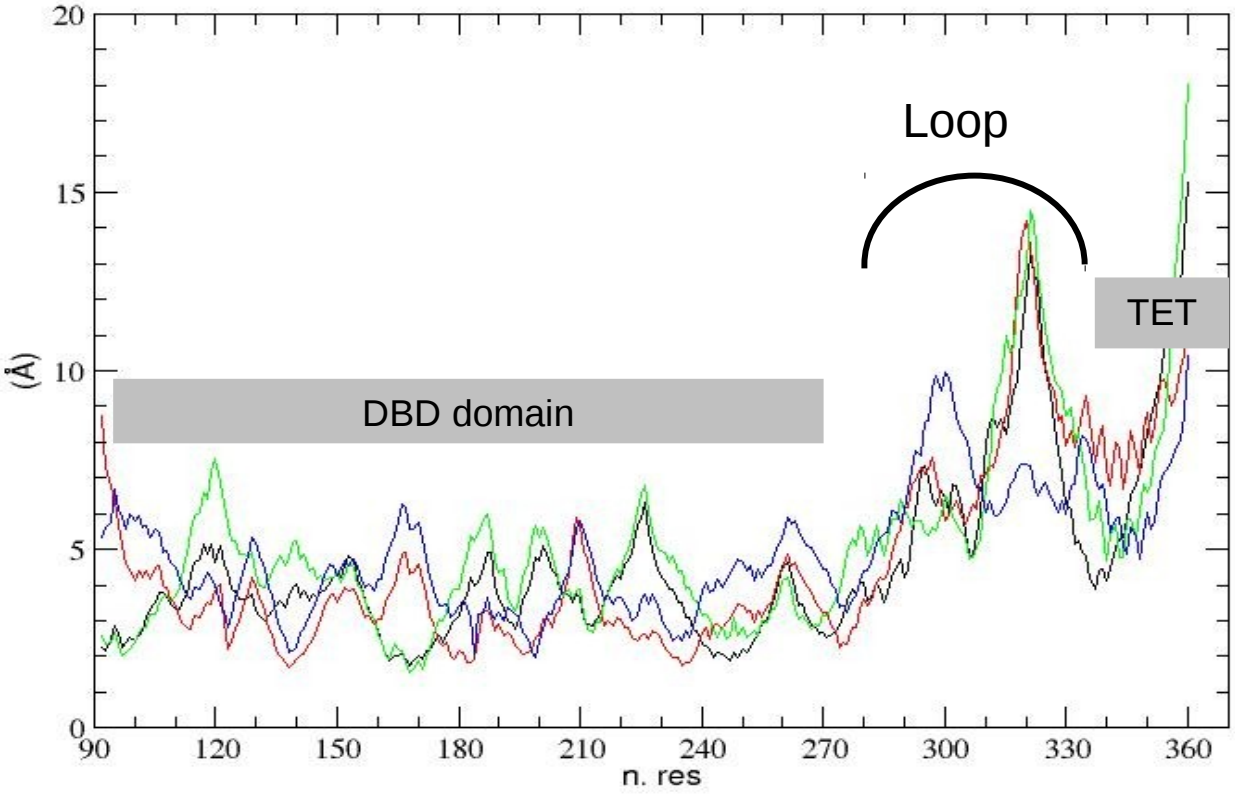
Expert Reviews in Molecular Medicine ©2003 Cambridge University Press

Measuring chain flexibility

System: p53 in water



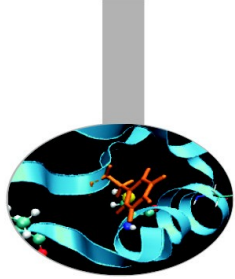
$$RMSF_i = \sqrt{\left\langle \left(r_i^{\min}(t) - \bar{r}_i \right)^2 \right\rangle}$$



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.

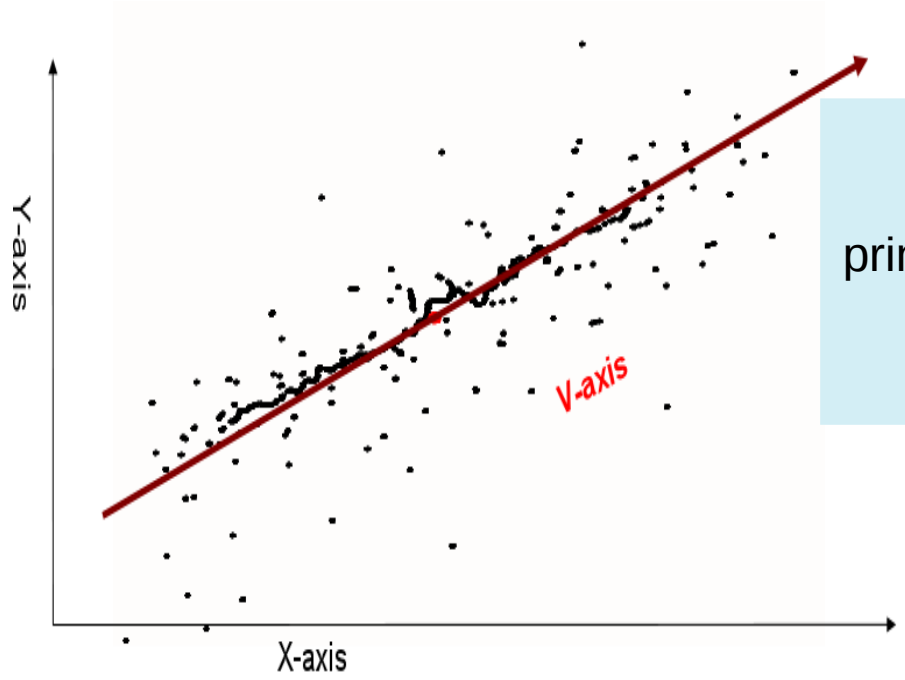
D'Abramo et al. Oncogene, accepted

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

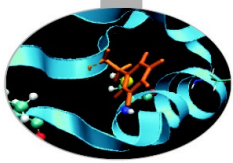


used for biological systems

principal modes ↔ biological function

Based on PCA

Essential Dynamics: workflow in GROMACS



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

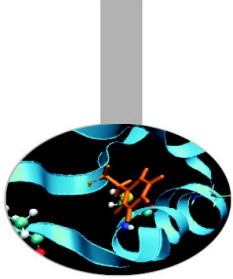
Calculate elements of the positional fluctuations covariance matrix of the Cα protein carbon atoms.

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

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

Sort eigenvector in descending eigenvalue index and determine principal componens

Eigenvalue equation



$$C\nu = \lambda\nu$$

Eigenvectors

Eigenvalues

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

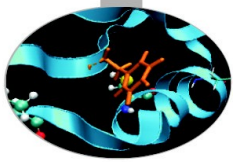
Moreover, it can be shown that σ^2 are numerically equivalent to calculated eigenvalues

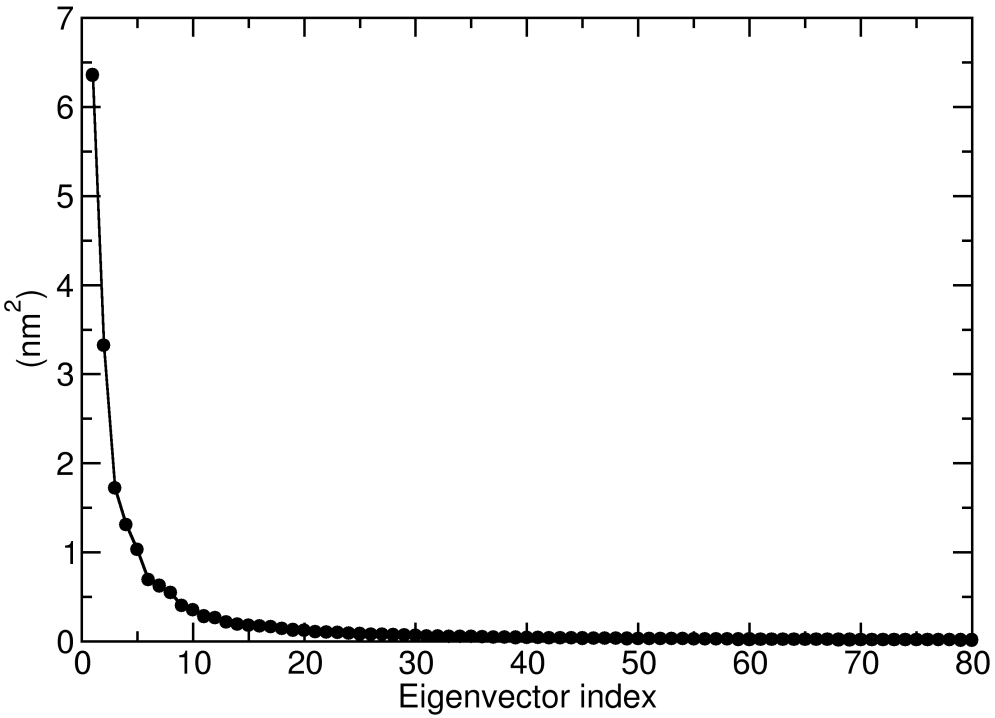
Eigenvectors, also called principal or essential modes.

R defines a transformation to a new coordinate system. The trajectory can be projected on the principal modes to give the principal components

pi(t): p(t) = R TM 1 2 (x(t) − hxi)

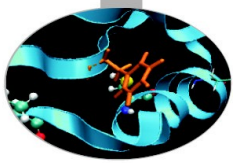
Essential Dynamics of Proteins

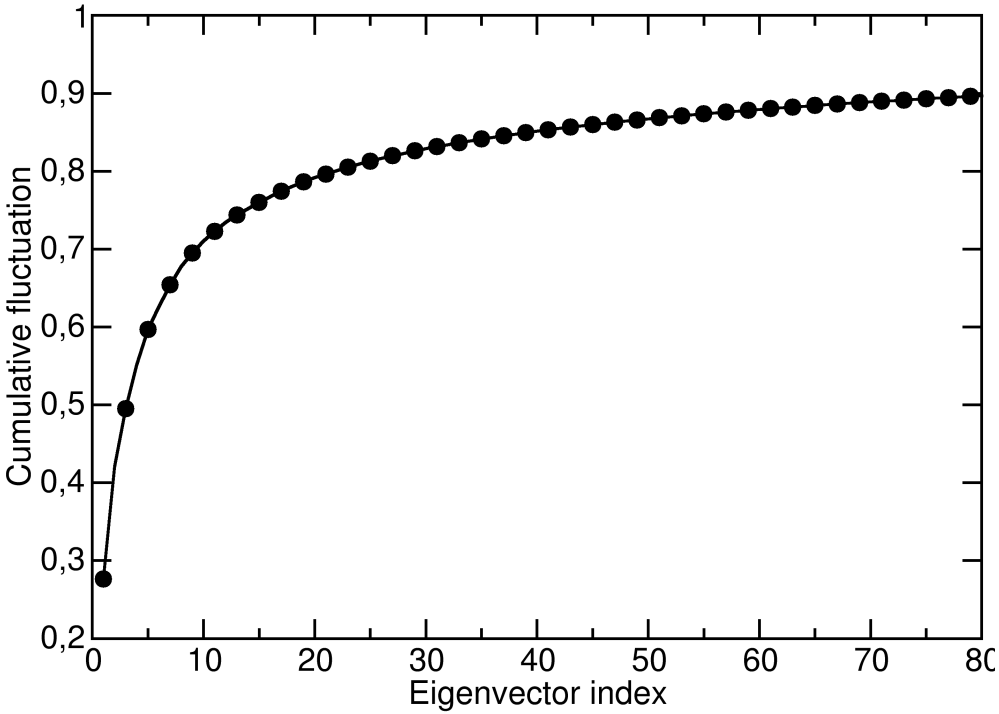




Eigenvalues are sorted in descending order: the first one corresponds to the maximum variance of the projected points. The corresponding eigenvectors are the best principal components of associated eigenvalues.

Essential Dynamics of Proteins

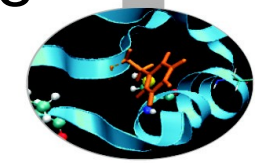




$$CF = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_n}{\sum_{i=1}^{N} \lambda_i}$$

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: general procedure



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

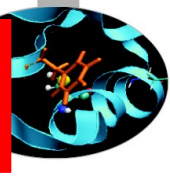
$$\Gamma_{ij} = \frac{1}{n}\sum_{h=1}^{n}\left(x_{hi} - \bar{x}_{\times i}\right) \times \left(x_{hj} - \bar{x}_{\times j}\right)$$

gmx 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

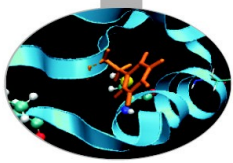


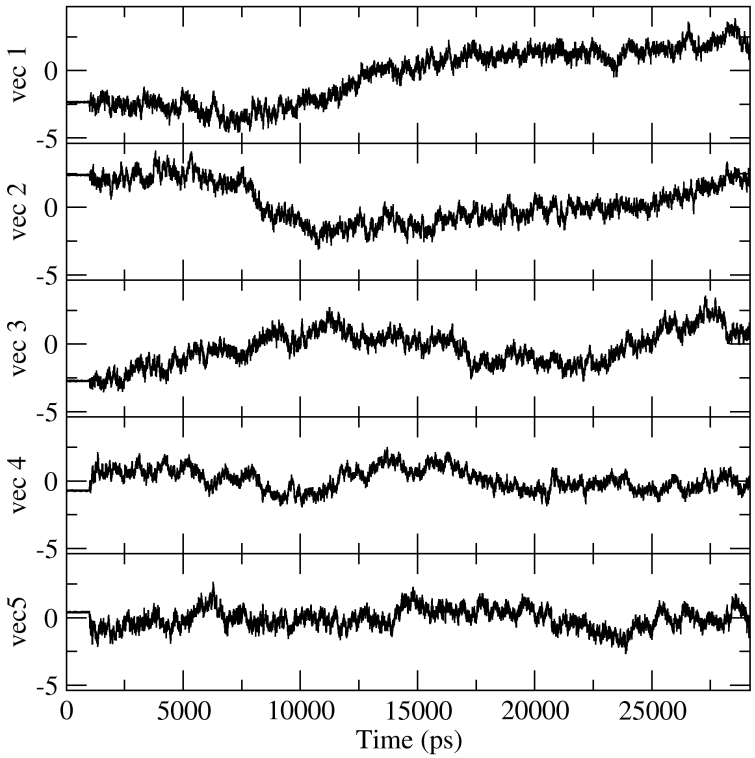
gmx anaeig –f trajectory.xtc –v eigenvec.trr –eig eigenval.xvg
–s reference.gro –b *start* –e *end* -first *eig-first* -last *eig-last*

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
-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

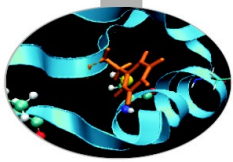


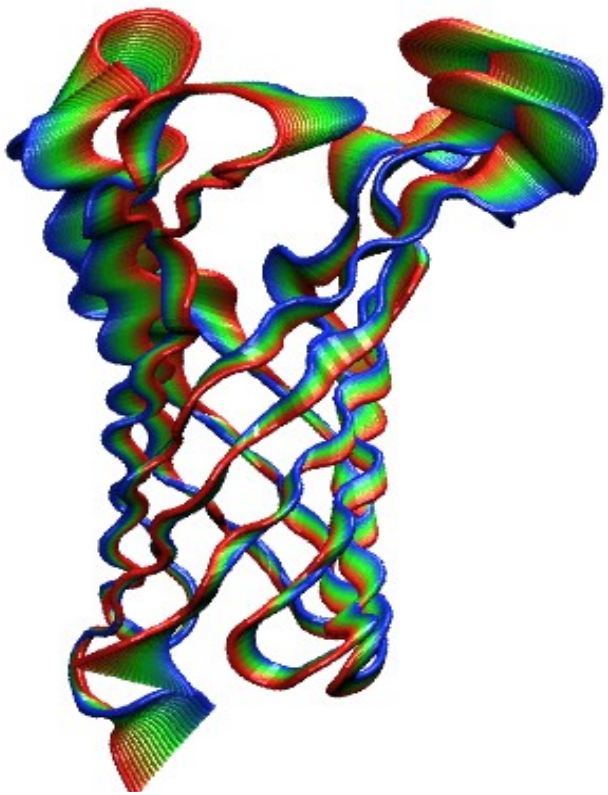


By default, 8 eigenvectors are considered for output using g_anaeig. This option can be set by using the flags `-first` and `-end`

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

g_anaeig: flag -extr

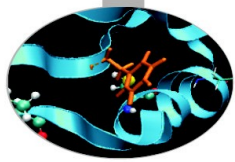


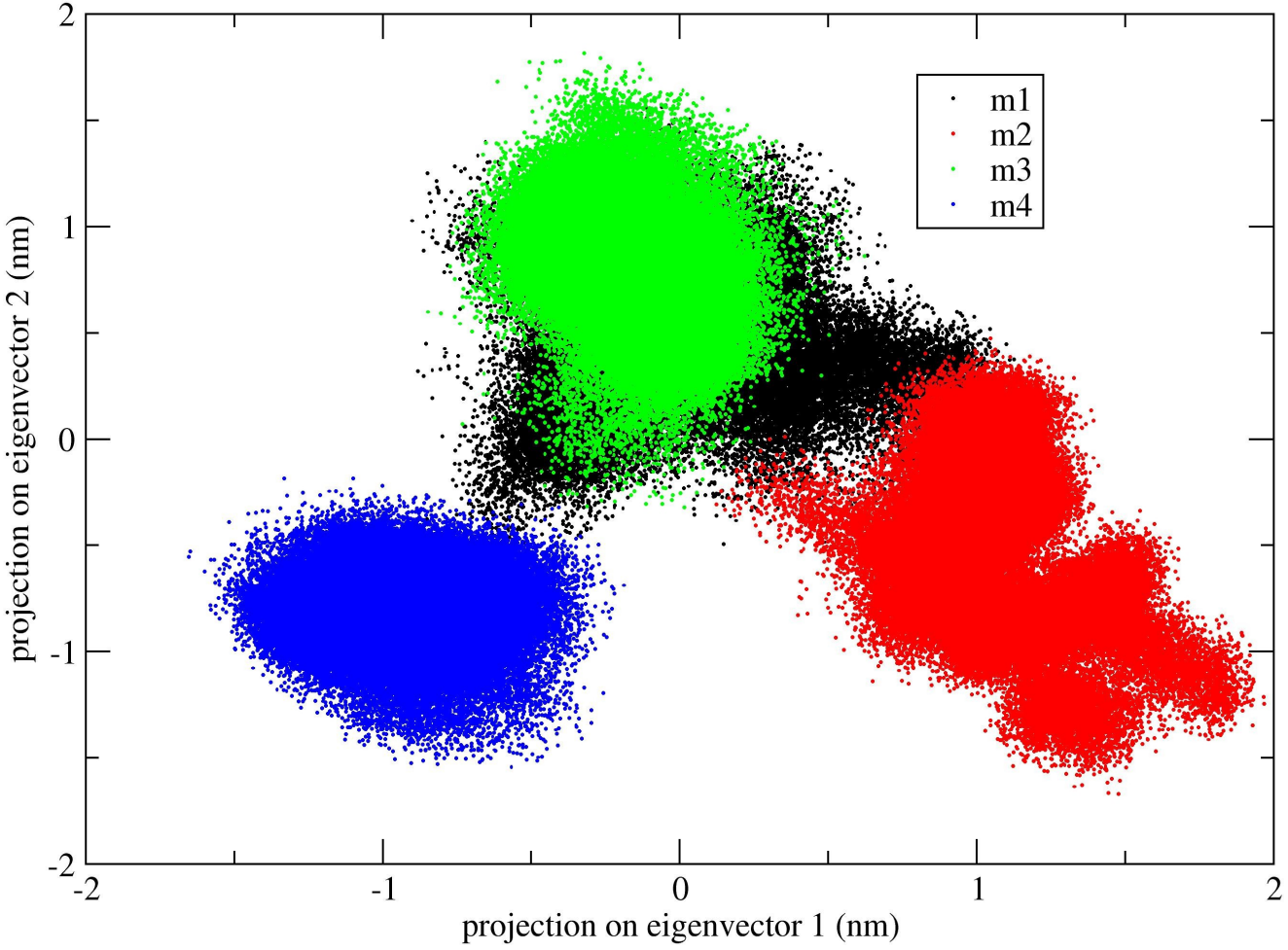


La dinamica essenziale ci aiuta a studiare i moti concertati tra gruppi di atomi all'interno di una struttura proteica. Nell'esempio si osserva un movimento concertato tra i loop extracellulari della porina OmpA lungo il primo autovettore.

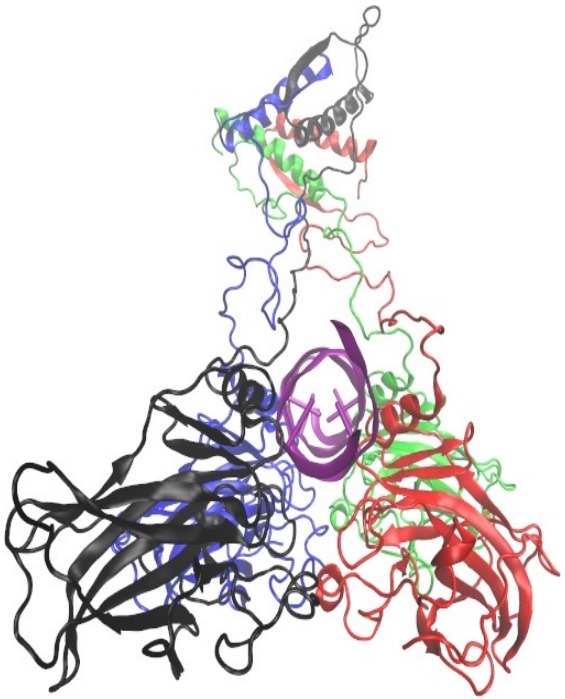
gmx anaeig -f trajectory.xtc -v eigenvec.trr -eig eigenval.xvg -s reference.gro
-extr extreme.pdb -first 1 -last 3 -nframes 50

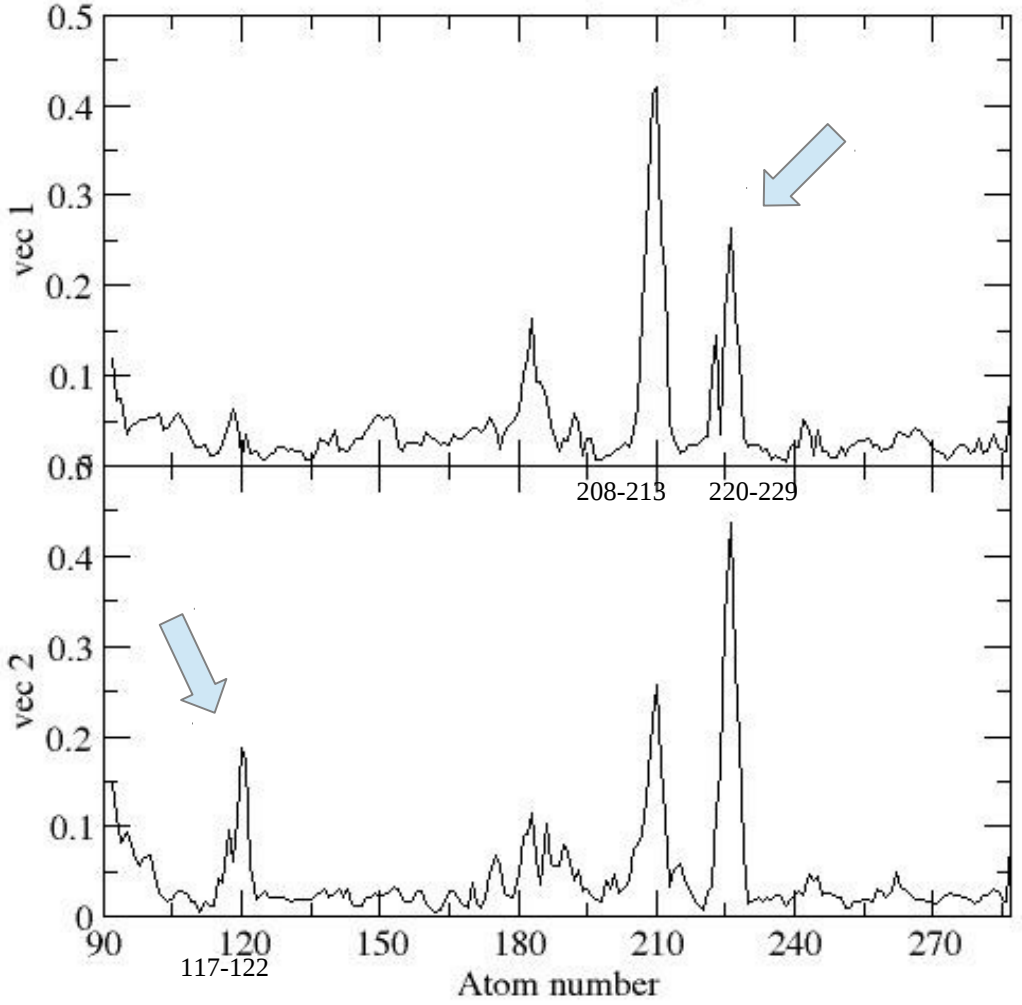
gmx anaeig: the -2d or -3d flag





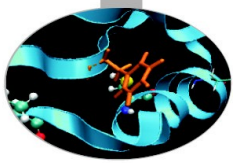
gmx anaeig: the –rmsf flag





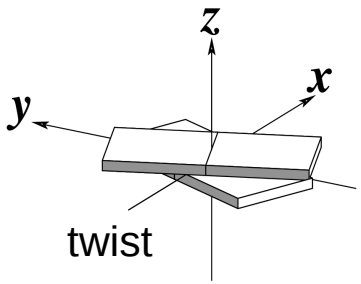
RMSF Analysis by means of principal components analysis.

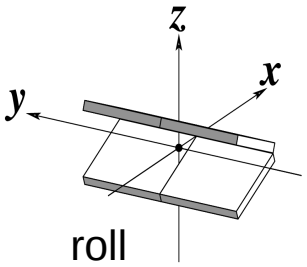
DNA structure

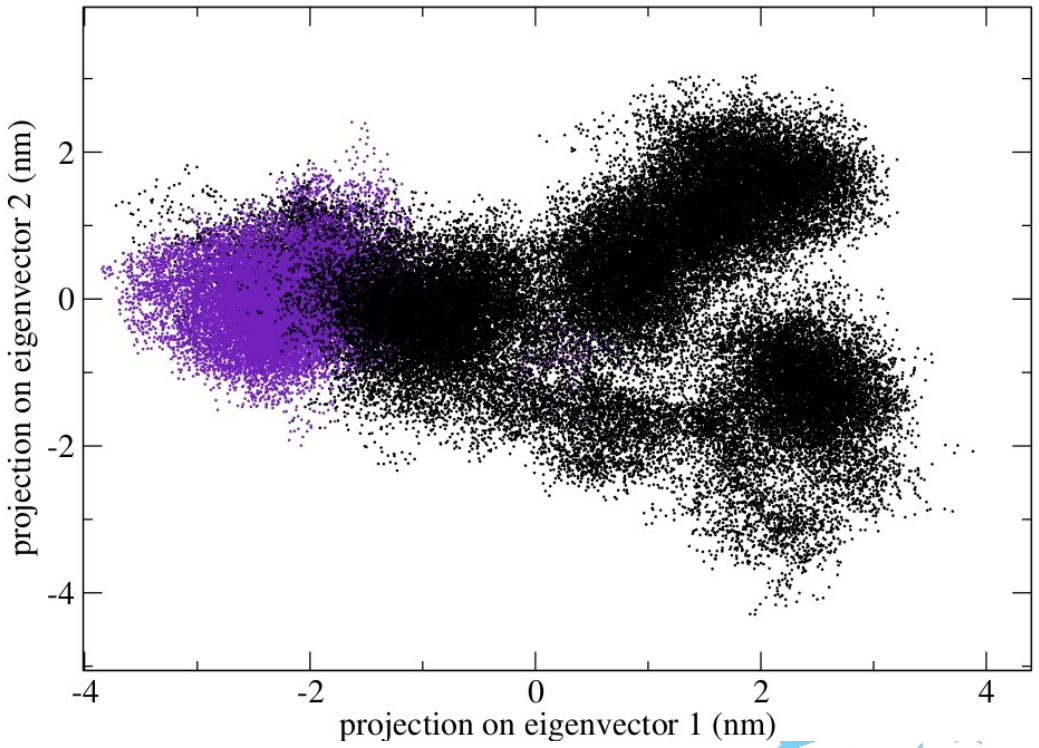


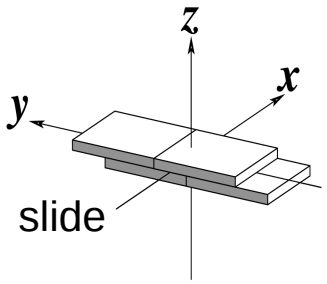
B DNA during the MD simulation

The ED analysis shows that the twist and roll distortion (observed in the X ray structure) of the central base pair is observed only when a specific region (well-described by the first eigenvector) of the tetramer conformational space is visited.

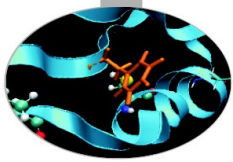


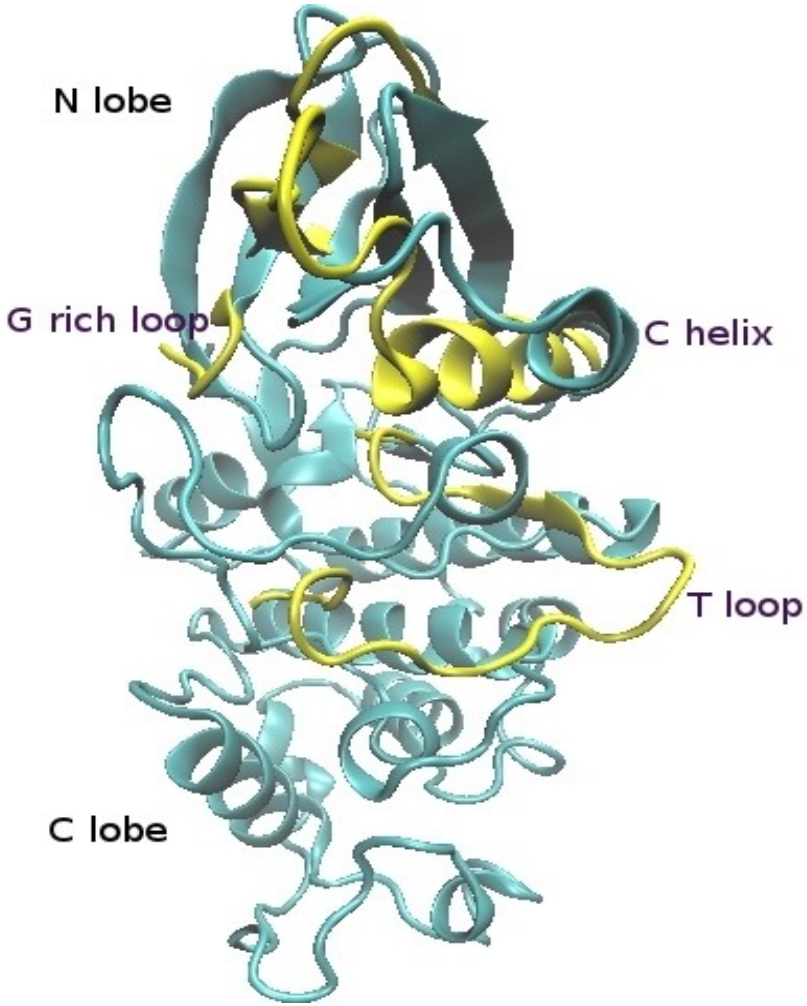


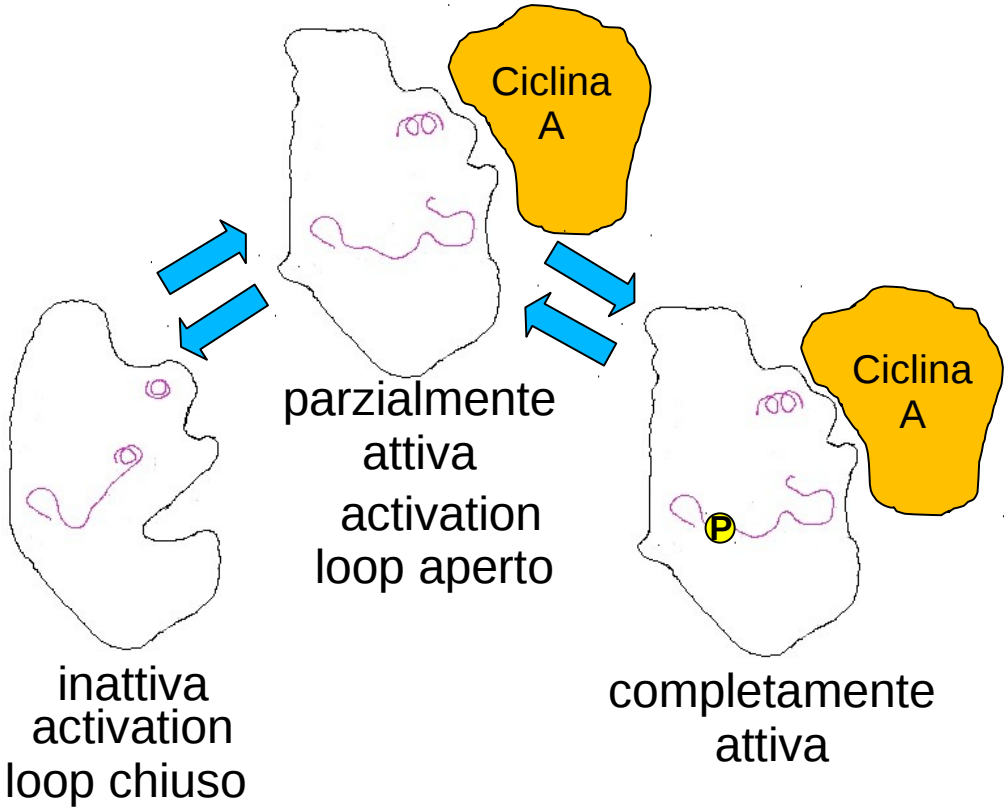




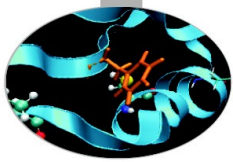
CDK2




N lobe
G rich loop
C helix
T loop
C lobe


Ciclina A
parzialmente attiva
activation loop aperto
Ciclina A
P
inattiva
activation
loop chiuso
completamente
attiva

Limite della simulazione MD classica : non vedo la transizione con due simulazioni MD di un μs ciascuna



ED sampling

- una simulazione di dinamica molecolare vincolata in uno spazio ridotto definito dal numero di autovettori essenziali

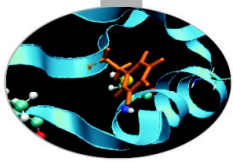
espansione
da una struttura di riferimento

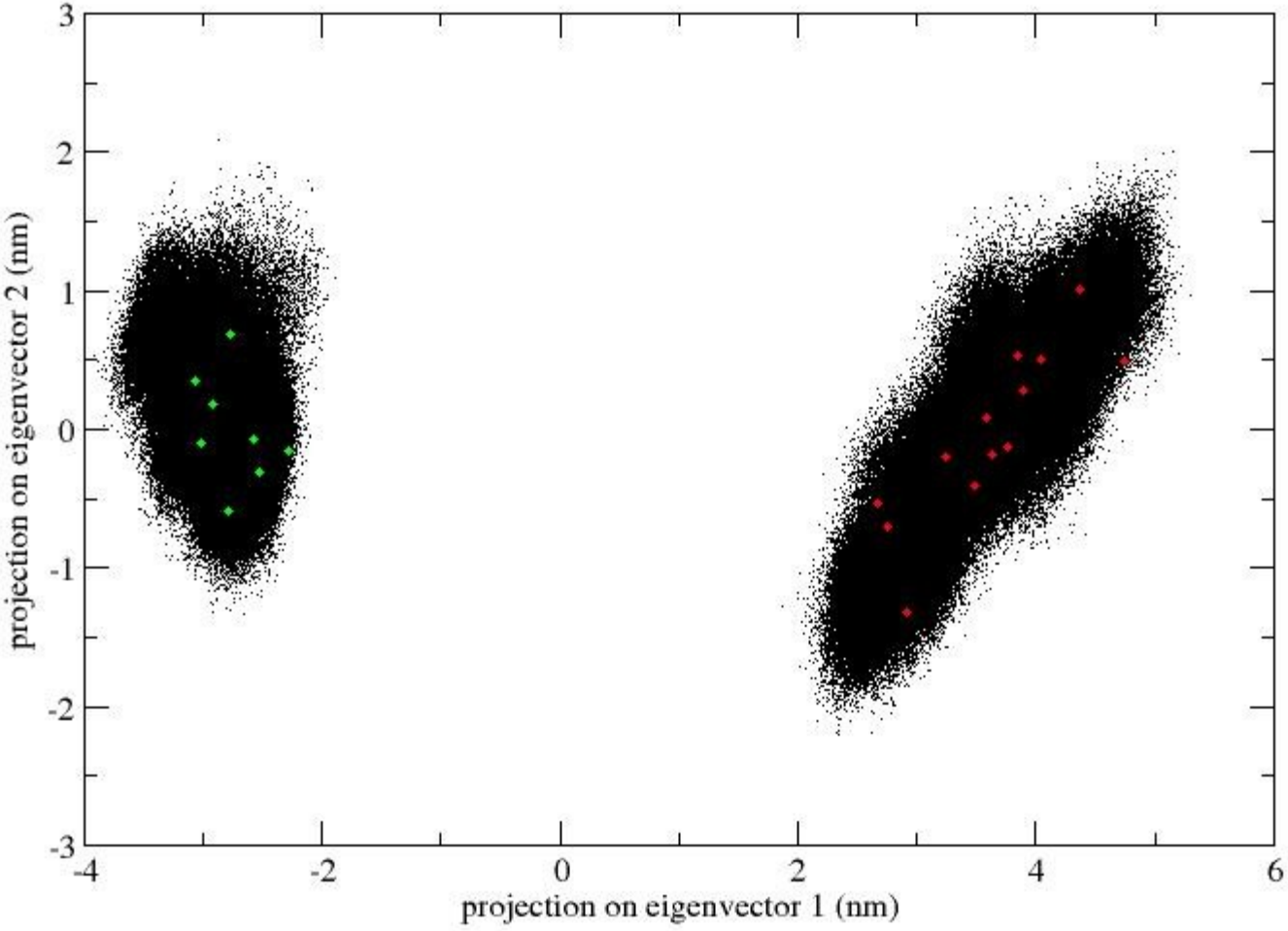
contrazione
decresce la distanza da una struttura di riferimento

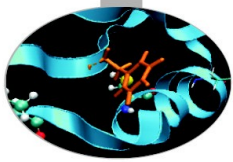
$\Delta\xi = \Delta\xi_d$ per Δt Per $r \leq r_0$

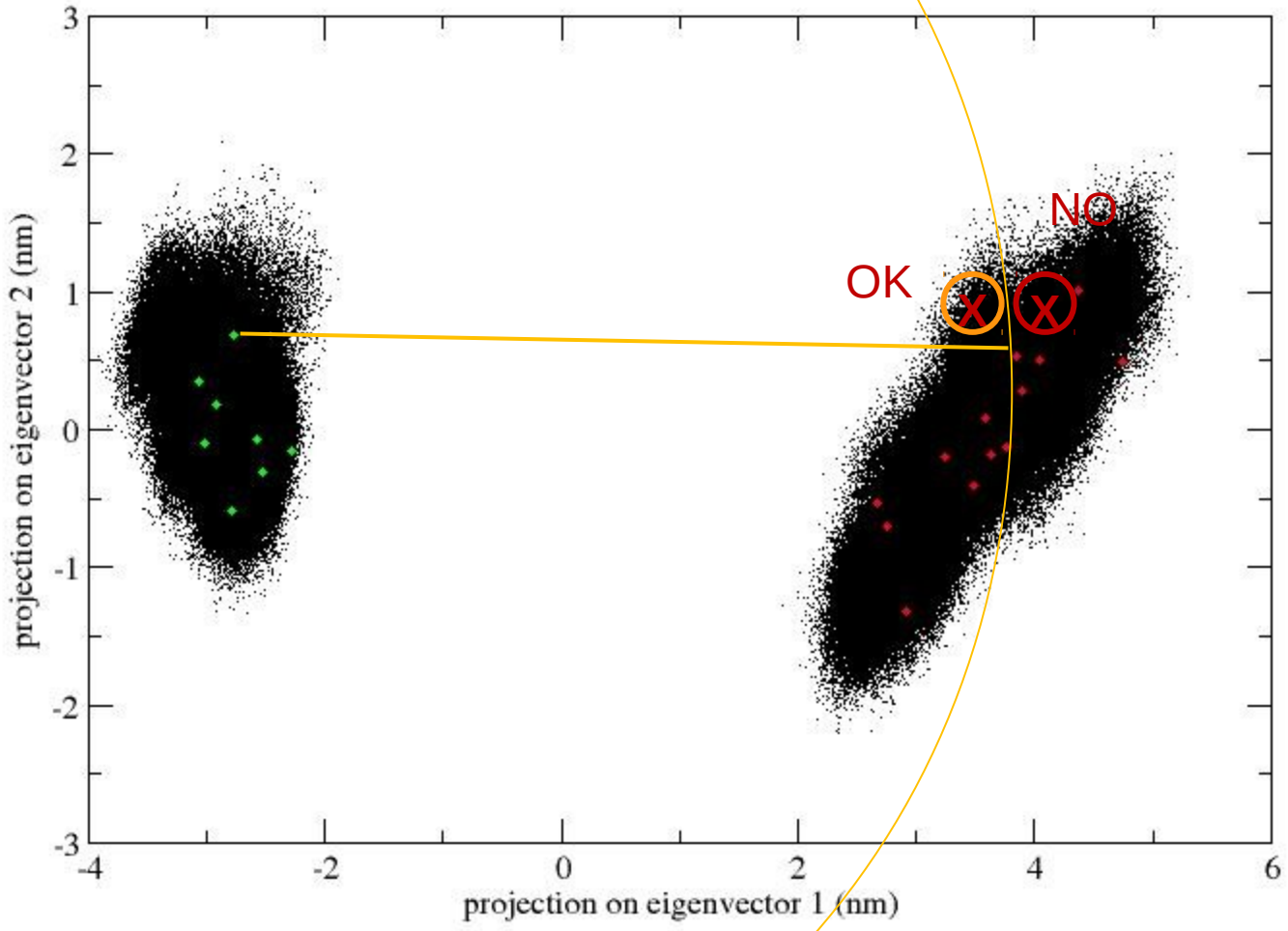
$\Delta\xi = \Delta\xi_d + \Delta\xi_c$ per Δt Per $r > r_0$

Amadei et al., J. Biomol. Struct&Dynamics, 13, 4, (1996), 615-625

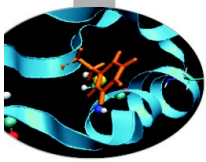


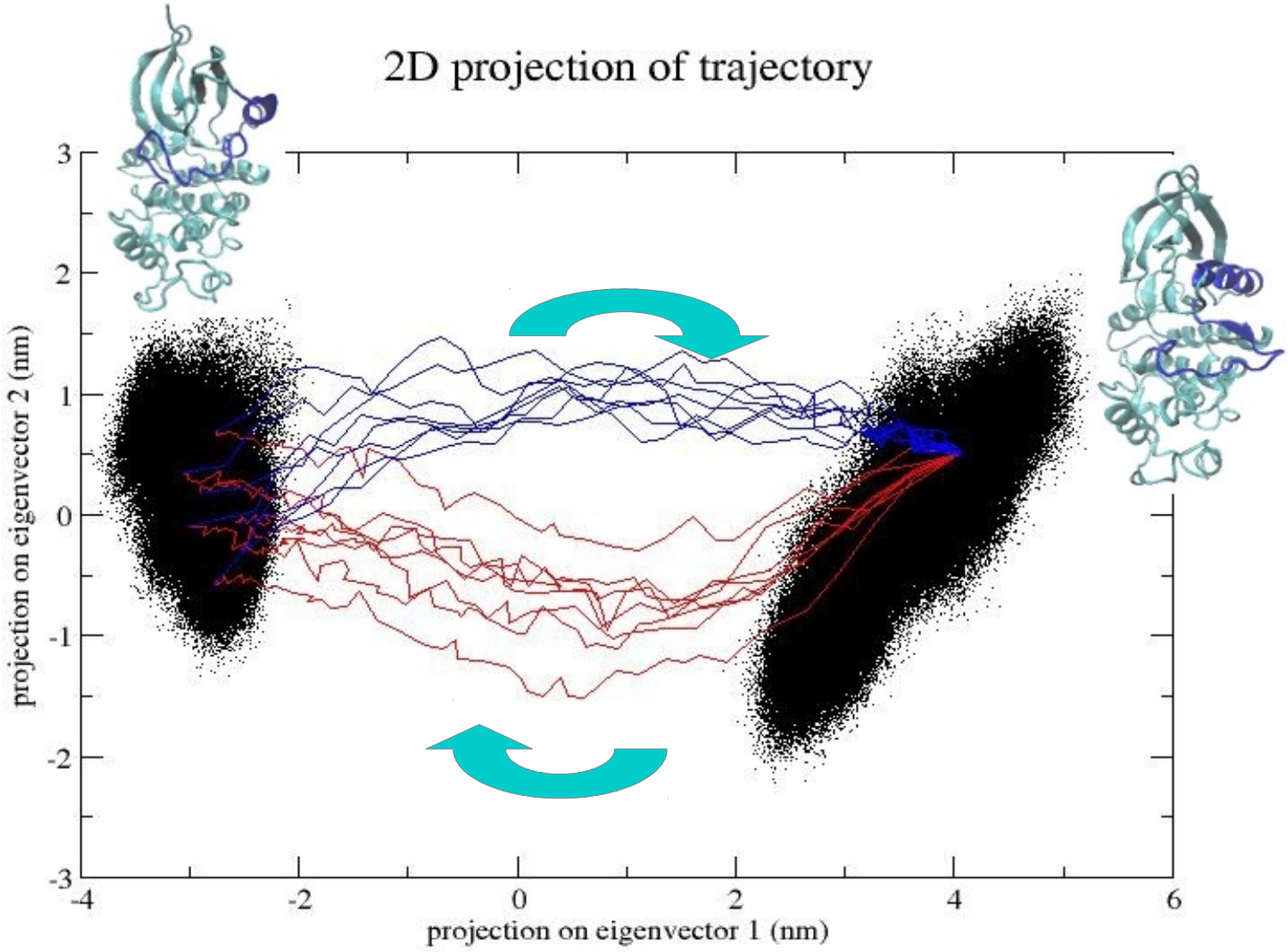

projection on eigenvector 2 (nm)
projection on eigenvector 1 (nm)
3
2
1
0
-1
-2
-3
-4
-2
0
2
4
6





IL SISTEMA è FORZATO A MUOVERSI LUNGO ALCUNE COORDINATE ESSENZIALI
Forziamo il sistema ad esplorare le nuovi regioni lungo queste coordinate

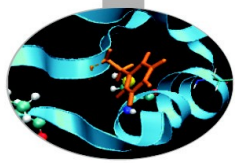




RMSD dalla struttura di riferimento; conservazione della struttura secondaria, conservazione dei legami H

Besker et al J. Biomol Struct Dyn. 2014 Dec;32(12):1929-35.

Essential Dynamics Tutorial: Principal Components Analysis of a small peptide



Input file on Eurora:
/gpfs/scratch/userinternal/agrottes/Corsi/September-2015/Tutorial3.tar.gz

Copiate in locale Tutorial3.tar.gz:

- scp username@login.eurora.cineca.it
- Run g_covar on file file.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