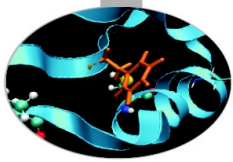
CINECA SCAI
SuperComputing Applications and Innovation



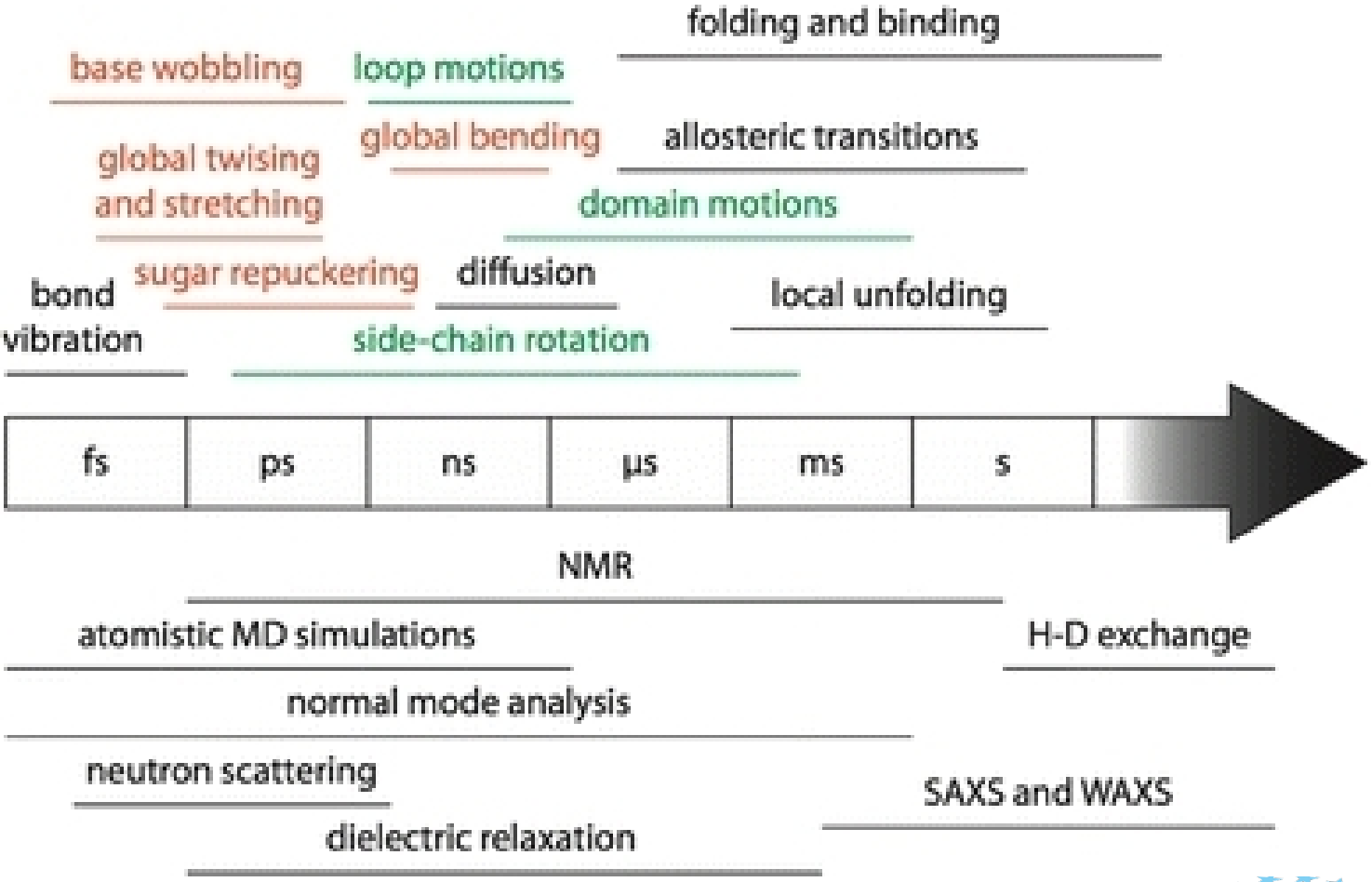
Essential dynamics sampling of proteins

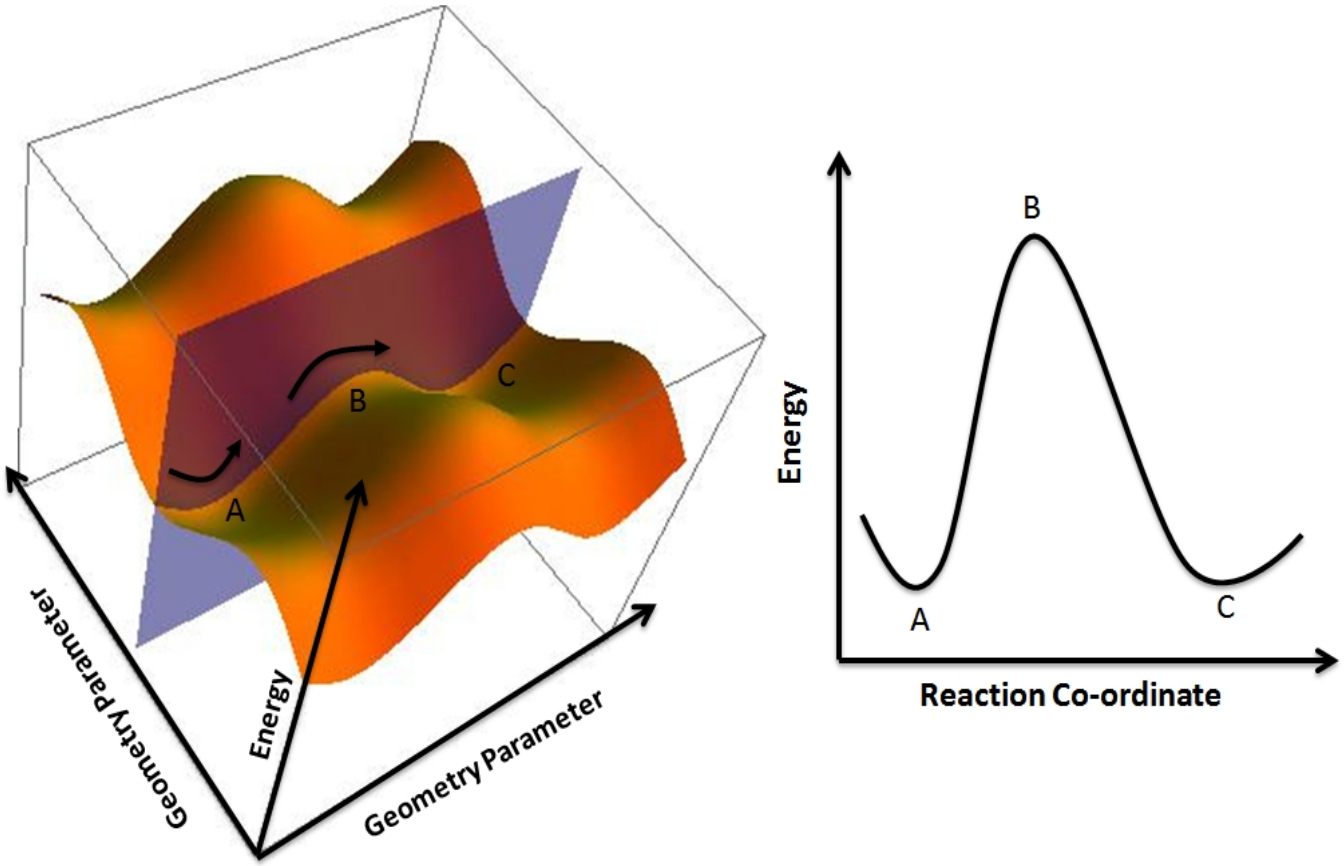
Tuorial 6
Neva Bešker

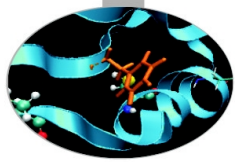
PRACE CINECA

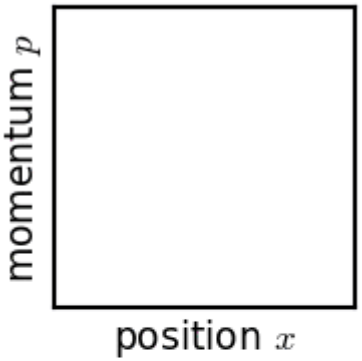
Relevant time scale

Why we need enhanced sampling?

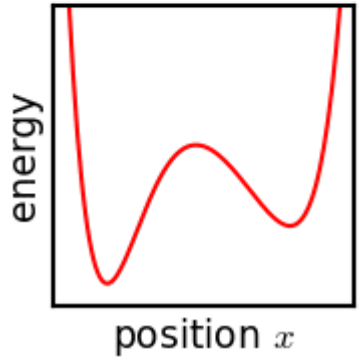



A
B
C
Geometry Parameter
Energy
Geometry Parameter
Energy
A
B
C
Reaction Co-ordinate





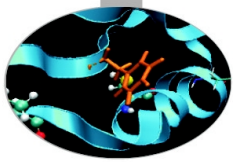
Interconvertion between basins is infrequent at the roomtemperature: kinetics and thermodynamics
Barriers are poorly sampled



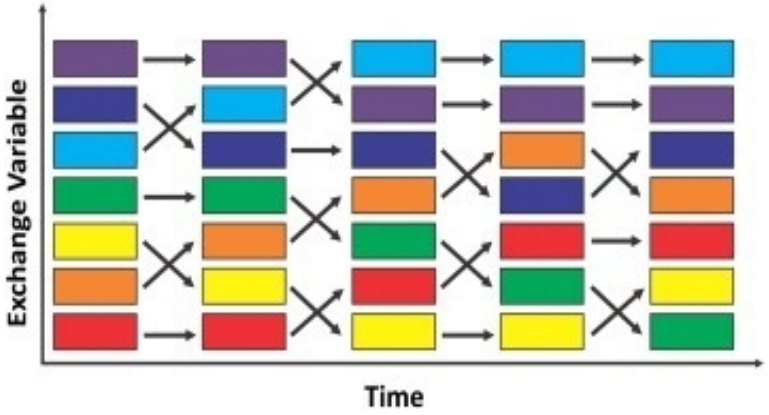
To overcome the limits a huge variety of sampling techniques has been developed

- Umbrella sampling
- Steered MD
- Replica Exchange
- Conformational flooding
- Metadynamics
- Essential dynamics

SAMPLING TECHNIQUES

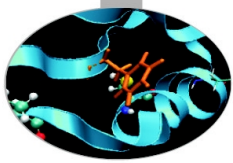


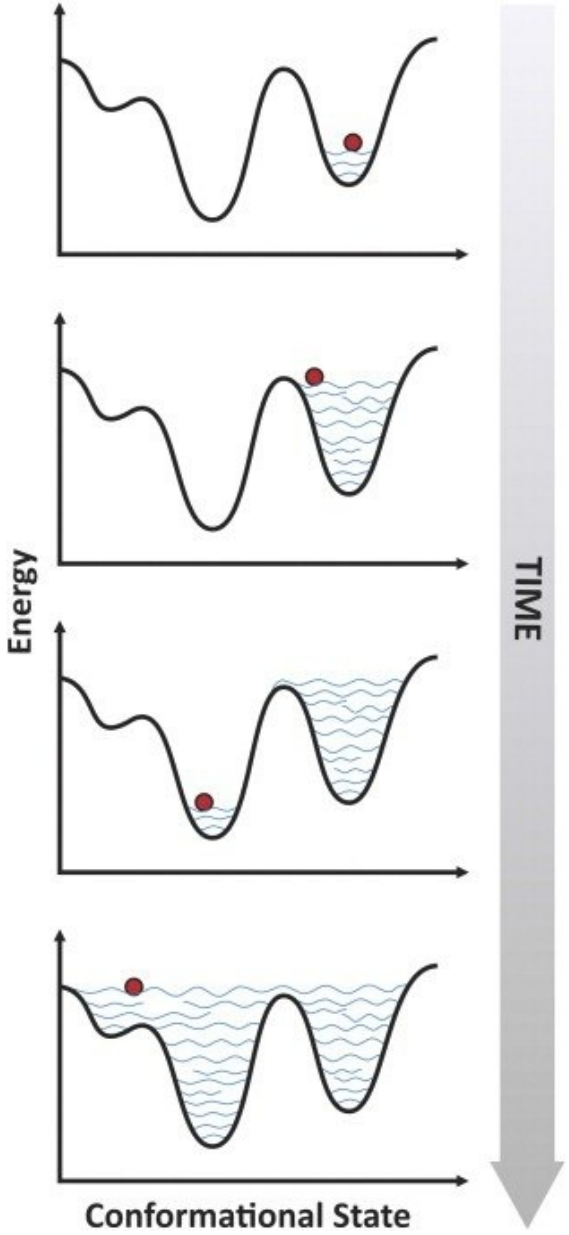
Replica-exchange Molecular Dynamics



Multiple indipendent MD siulations are run with different values of a specific exchange variable (Temperature). At certain time intervals system states exchanged between neighbouring simulations based on Monte Carlo acceptance scoring algorithm.
The effect is to overcome energy barriers on the potential energy surface

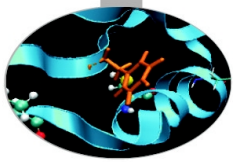
Metadynamics





The aim is to add "memory" into the sampling process, thus preventing oversampling of local energy minima. Once a state has been sampled, a positive Gaussian potential is added to the real energy landscape to discourage the re-sampling of previously visited states.

This can be thought of as “filling the free energy wells with computational sand”



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

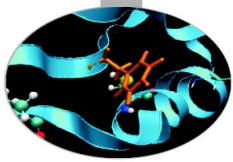
expansion procedure increase the distance from a reference structure

contraction procedure decrease the distance from a reference structure

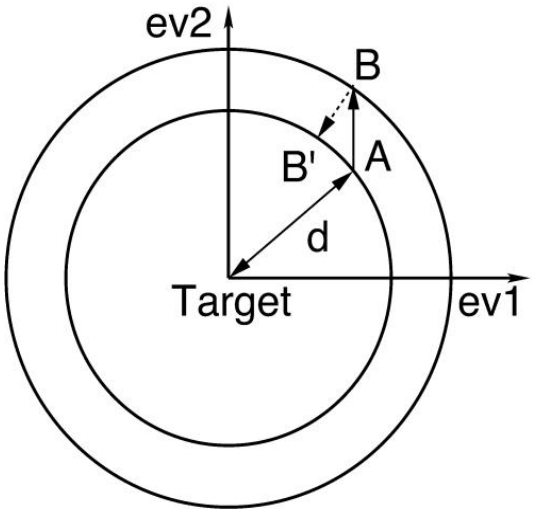
$\Delta\xi = \Delta\xi d$ per Δt Per $r \leq r0$

$\Delta\xi = \Delta\xi d + \Delta\xi c$ per Δt Per $r > r0$

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



For each step a regular MD simulation is performed and the distance between the current structure and the refference structure is calculated



The step is

→ ACCEPTED

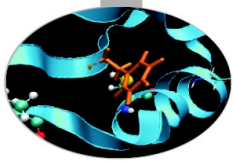
If the distance doesn't decrease (expansion)

→ IF NOT ACCEPTED

Coordinates and velocities are projected radially onto the hypesphere of the chosen subspace

No additional forces are added!

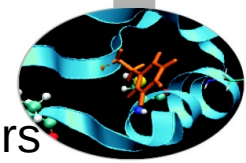
Essential Dynamics sampling



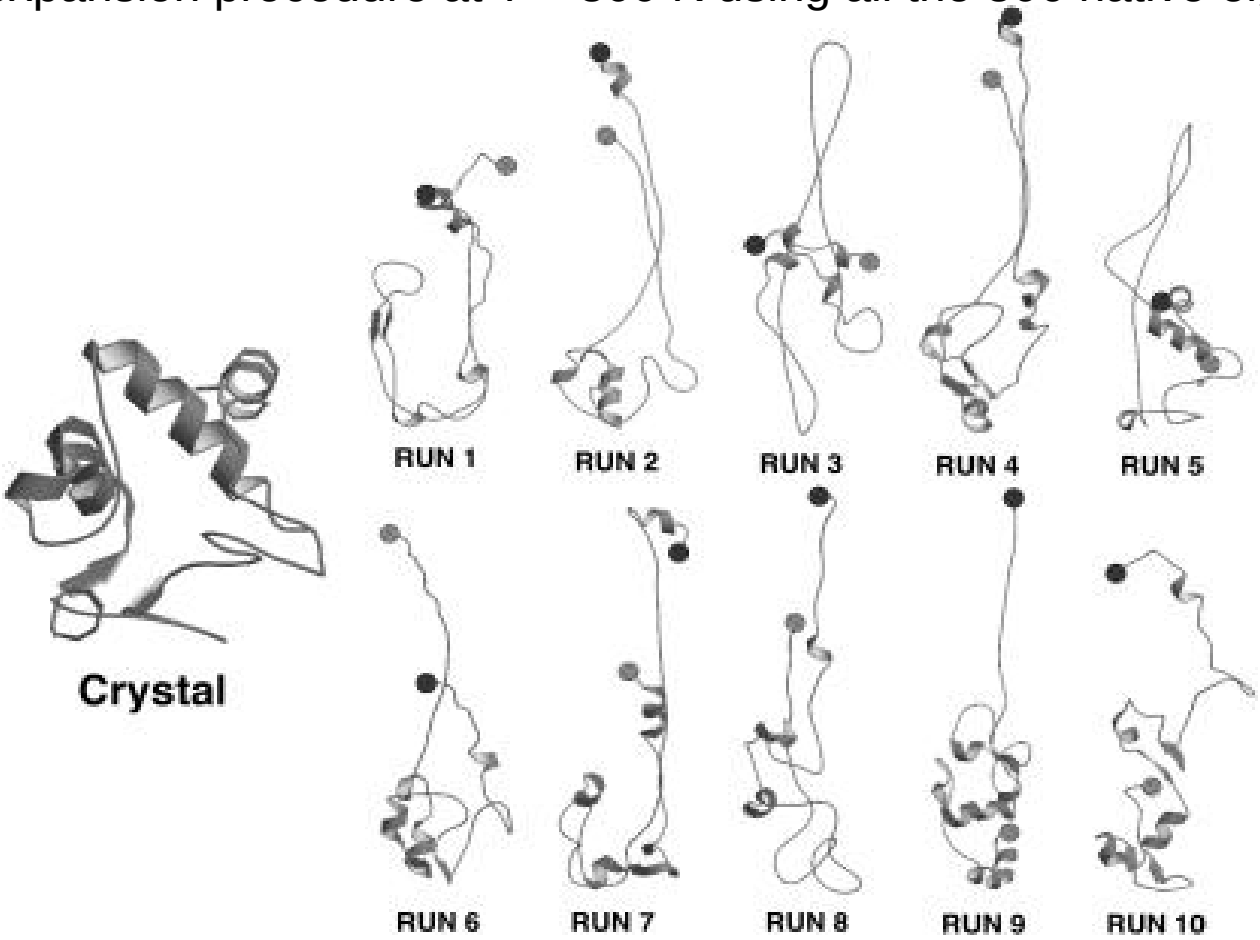
Constrained dynamics simulations in the reduced space defined by a number of essential eigenvectors

- Folding / unfolding transitions in peptides and proteins
- Conformational transition in proteins

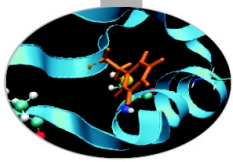
Folding cytochrome C

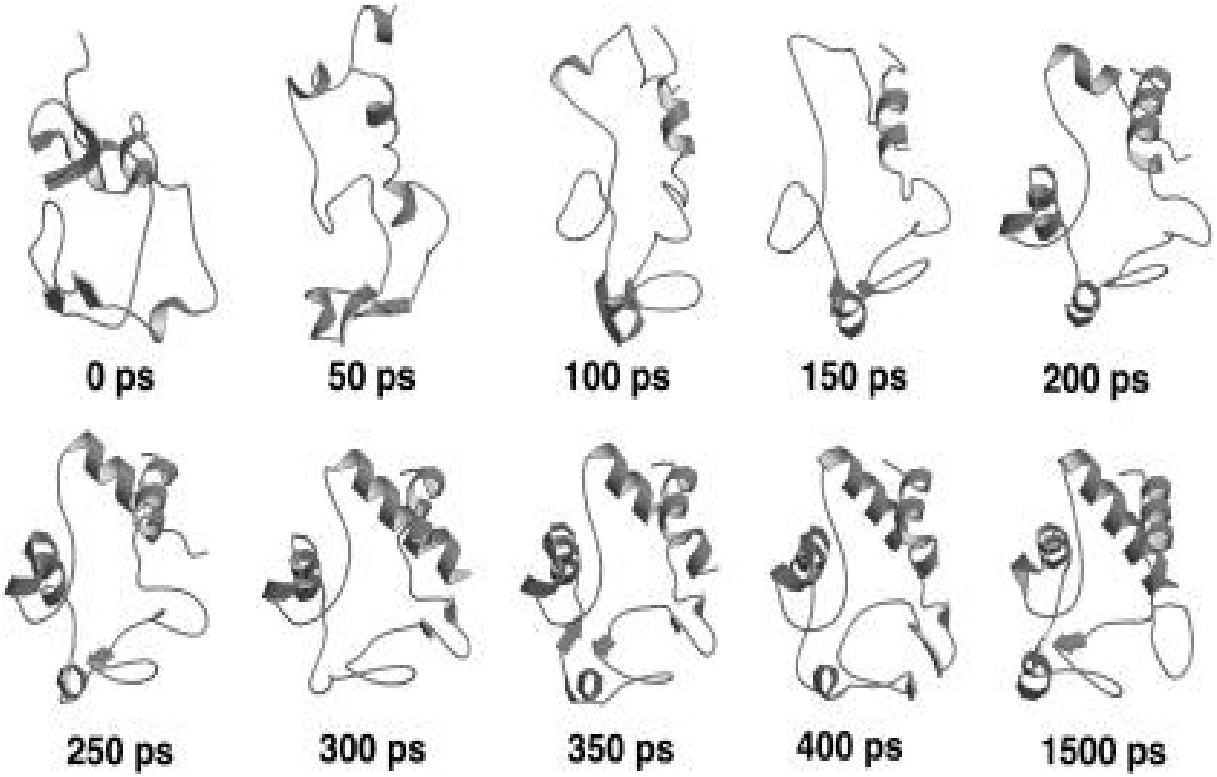


EDS expansion procedure at T = 300 K using all the 306 native eigenvectors



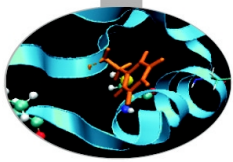
eigenvectors 1–100, 101–200, and 201–306





Ribbon diagrams of sequential snapshots along the refolding trajectory using the SET with 201-306 eigenvectors.

Daidone et al. Biophys. J. 2003, 85:2865-2871
Narzi et al. JCTC 2008; 4:1940-1948

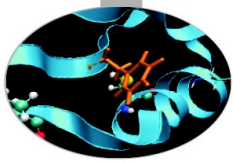


Successful folding only with the eigenvectors with lowest eigenvalues, representing the most rigid quasi-constraint motions.

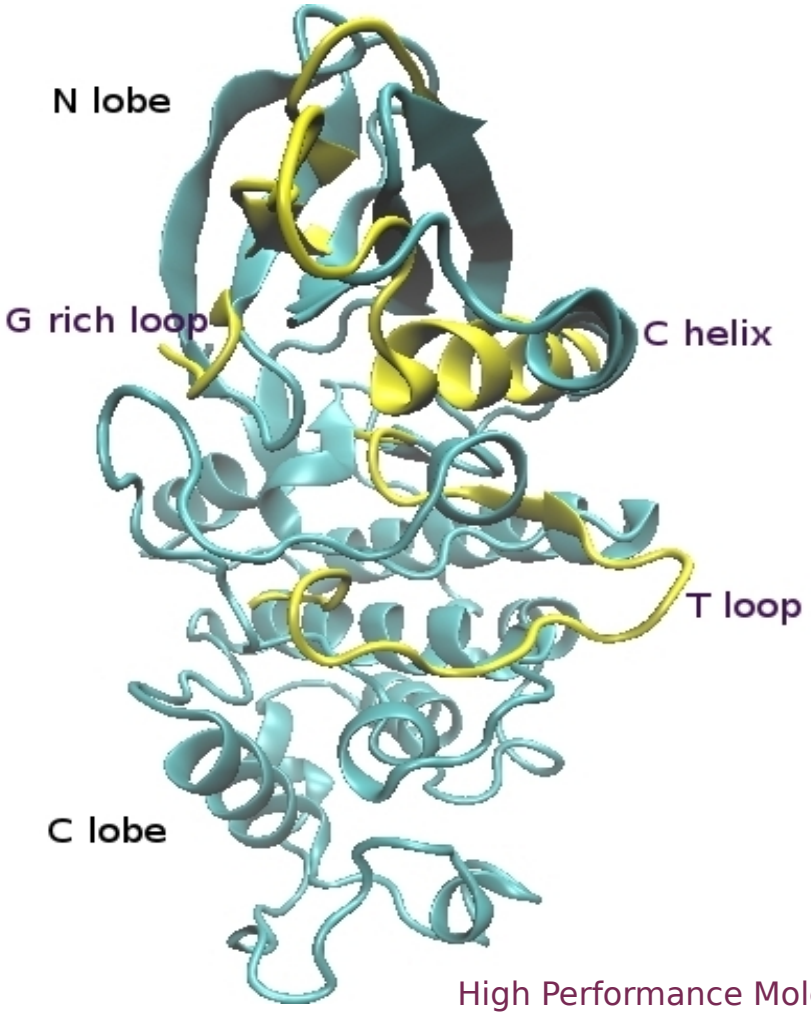
If the essential eigenvectors, the ones accounting for most of the variance, are used, folding is not successful. These results clearly show that the eigenvectors with lowest eigenvalues contain the main mechanical information necessary to drive the folding process, while the essential eigenvectors represent the large concerted motions which can occur without folding/unfolding the protein.

EDS to model conformational transition

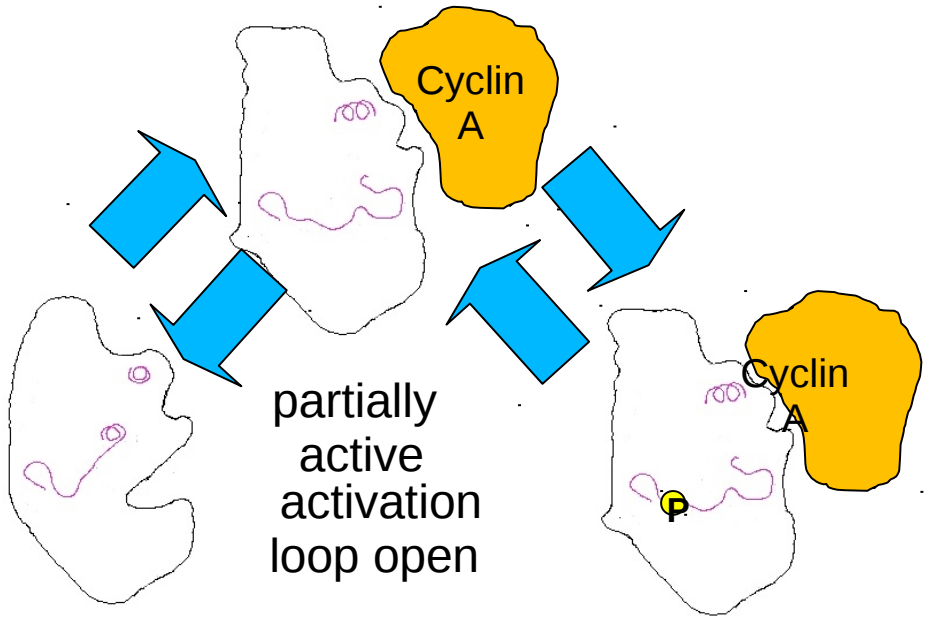
CDK2 example



Limit of the classic MD simulations:



no transition in two MD simulations lasting 1 μs



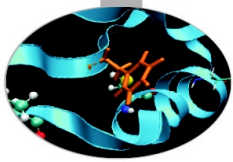
Inactive form activation loop closed

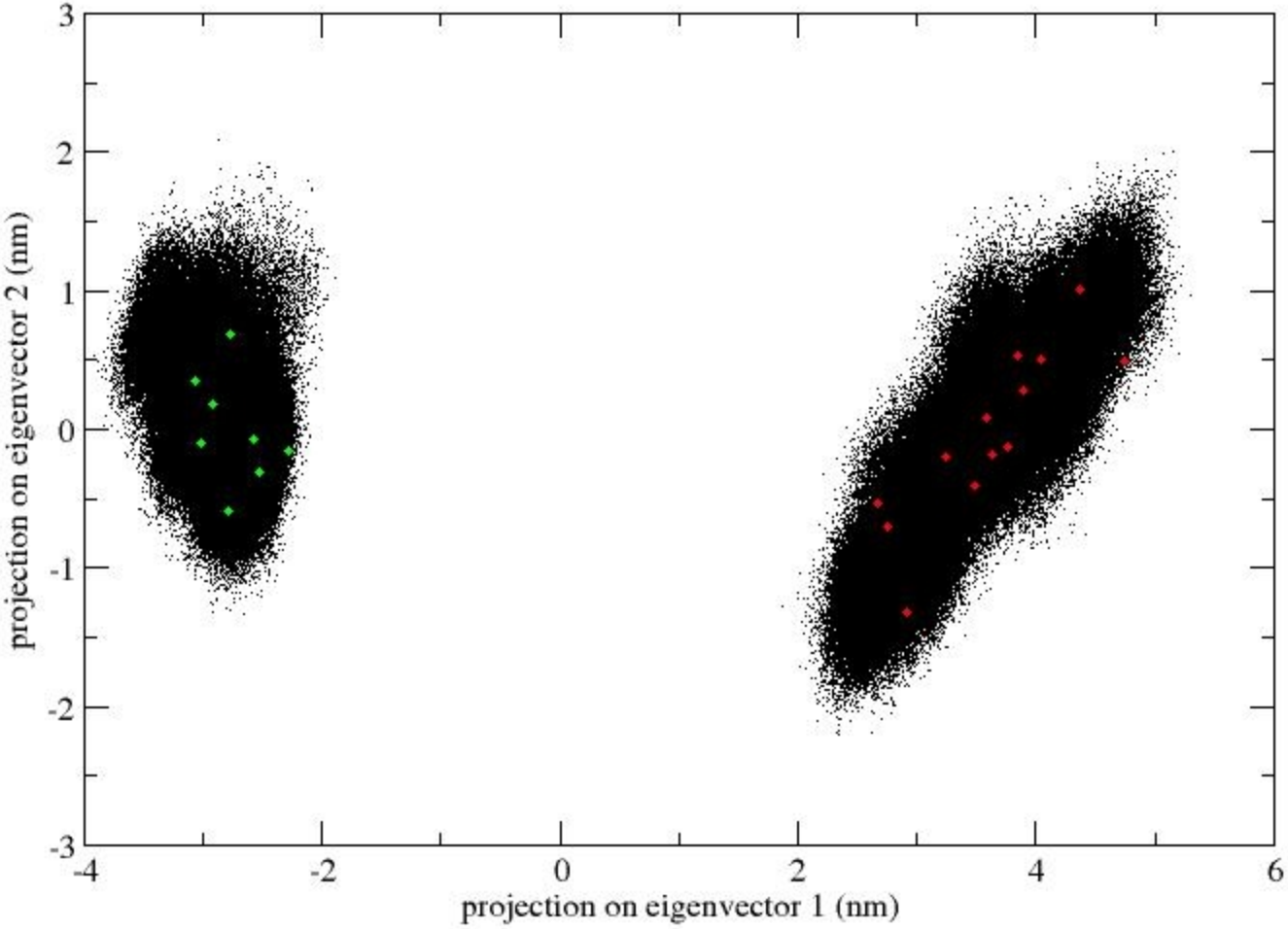
Different orientation of Cα helix

fully active

PRACE

CINECA

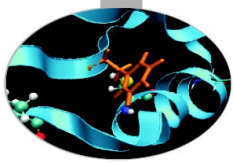


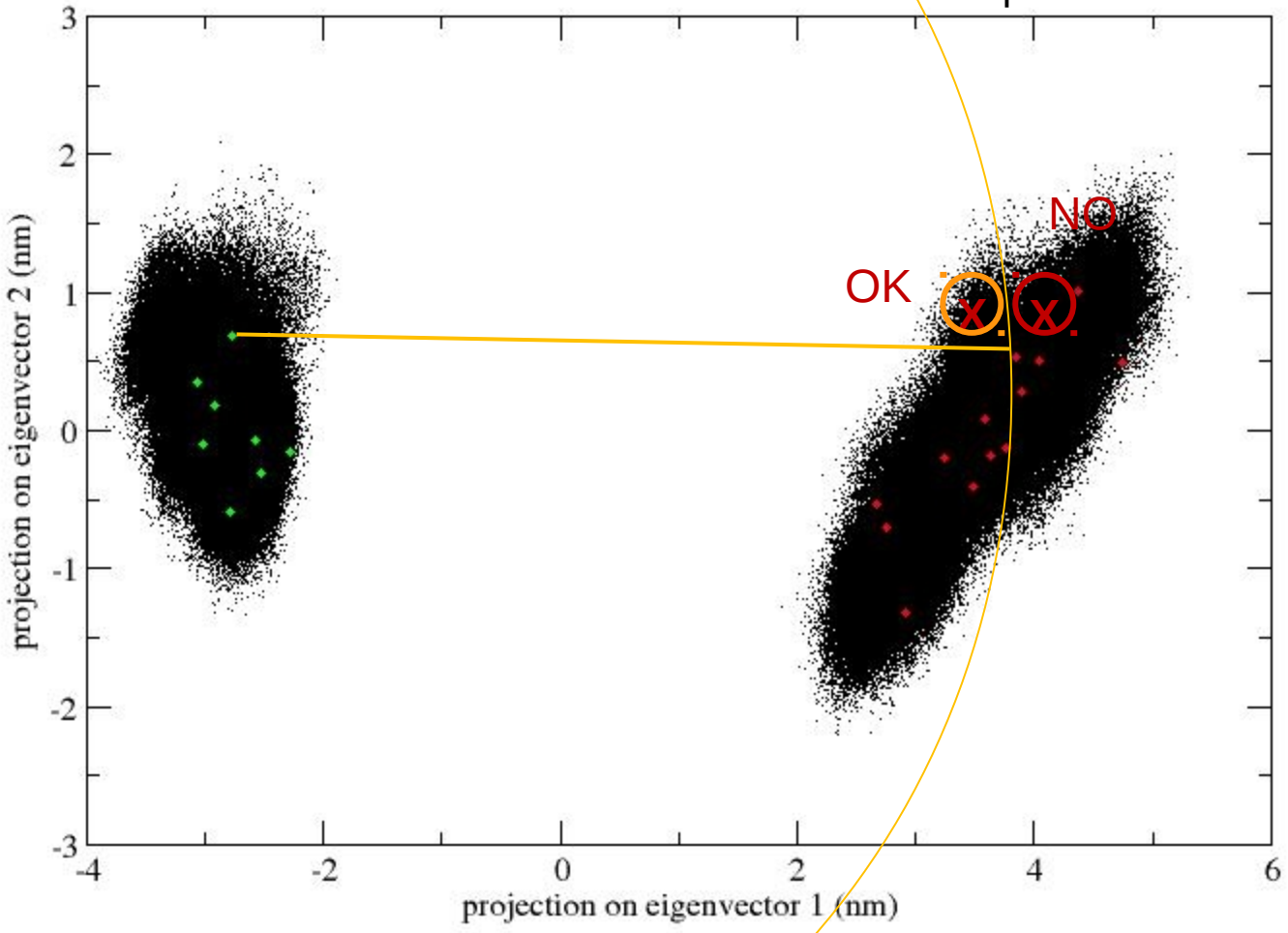


the two trajectories were concatenated, covariance matrices of positional fluctuations (C-alpha only) were built and diagonalized.

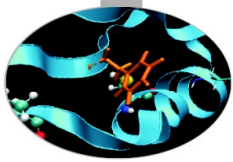
which eigenvector – obtained by means of principal component analysis – well discriminate between the two protein conformational states ?

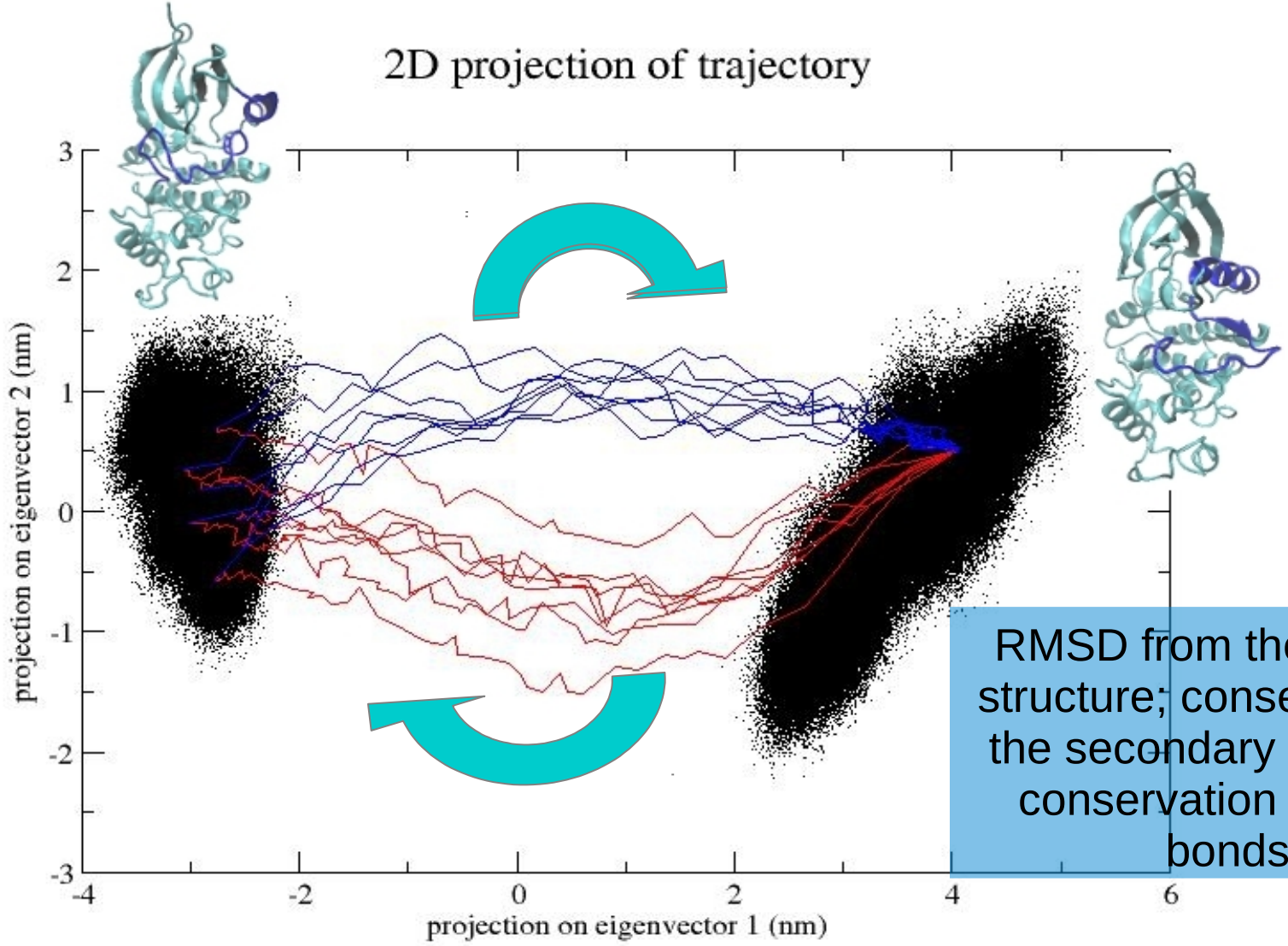
Hypersphere with r=the distance between two structures on the essential subspace





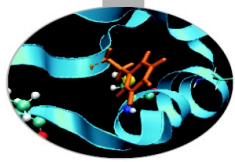
THE SYSTEM IS CONSTRAINED TO MOVE TOWARDS A REFERENCE POINT (the origin of the hypersphere) IN THE ESSENTIAL SPACE





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.



CONCLUSIONS

PCA is not based on any a priori information

The method allow to provide (physically consistent) low dimension subspace

Possibility to collect a huge amount of the new data (sampling) in order to evaluate the termodynamics and kinetics of conformational transitions

The EDS method Is still undergoing new developments...